International Education Journal

IEJ

Volume 1 Number 1
December 1999

Published by

Shannon Research Press
Adelaide, South Australia
ISSN 1443-1475
http://iej.cjb.net

International Education Journal

Volume 1, Number 1, December 1999

Welcome to the International Education Journal. This is a broadly based journal encompassing research and review articles. Education is interpreted in a wide manner and includes human development, learning, school education, formal and informal education, tertiary and vocational education, industry training and lifelong learning.

Online - While there will be a hard copy version of the journal it is intended that it will be accessed in almost all instances via the web.

Peer Review - All articles appearing in this journal have been subjected to a peer-review process.

Published by:
Shannon Research Press,
13 Endeavour Court,
North Haven, SA 5018,
Australia

ISSN 1443-1475

Designed by Katherine Dix

Printed in Adelaide
December 1999

CONTENTS

FREE Online Subscription
Subscription is available as a free service in online format only. Either complete the subscription form at: http://iej.cjb.net or email the onlineeditor@iej.cjb.net. The reader will be notified of the latest issue and contents.

Publication Frequency
It is intended that the journal be published three times per year. However given the electronic characteristics of online publishing it is not necessary to wait until a full journal is assembled prior to publishing. Articles once accepted in appropriate form may be published without delay. It should be noted however that in electronic format pagination may vary according to the parameters of the downloading computer. Alternatively the journal article can be downloaded in PDF format, in which case the original pagination will be preserved.

Aims of IEJ

The aim of the International Education Journal is to publish articles that possess one or more of the following characteristics or qualities.

1. The presentation of substantive findings which are of importance for policy and practice beyond the country in which the research was conducted.
2. The integration of education with academic disciplines such as anthropology, demography, economics, history, law, linguistics, philosophy, political science, psychology and sociology, or examines educational issues from the perspective of the disciplines or investigates issues at the interface between education and one or more of these disciplines.
3. The examination of educational issues from a cross-cultural or indigenous people's perspective.
4. The evaluation of educational policy or programs or the use of information technology of cross-national interest and significance.
5. The employment of advanced research methods and measurement procedures which are clearly explained.
6. The presentation of empirically or analytically based investigations of theory, models or conceptual frame works in the field of education.
7. The syntheses of research findings from comparative and cross-national studies in education.

Invitation

Contributors - Authors are invited to submit material to this journal. As a general guide articles should be not more than 5000 words. The journal may publish longer works such as theses or disserations as occassional mongraphs. In every instance material must be received in publisher ready format. Full details of publication style and other requirements are set out under the heading Author Information.

Publishers - If you wish to have books or educational software reviewed and published in the International Education Journal please contact the editors.

Reviewers - If you are interested in contributing to the journal by writing a review article (500-1000 words) please contact the authors indicating your areas of interest. We look forward to hearing from you.

Author Information

IEJ welcomes practical and research manuscripts that focus on educational issues, provide a clear purpose and depth of discussion, and are presented in a straight-forward style accessible to an international audience.

The initial submission:

- should be prepared using the downloadable Microsoft WORD97 template document (iejbase.doc), complete with layout styles, images and tables correctly positioned in the text
- images that are imported or inserted into the text from an external source as figures should be included as separate graphic files (either GIF or JPG format)
- should be forwarded electronically, either by:

  e-mail to: onlineeditor@iej.cjb.net

  or,

  on a 1.4MB 3.5 inch floppy diskette to: IEJ Online Editor, 13 Endeavour Court, North Haven, SA 5018.

The editors reserve the right to return poorly edited or improperly formatted manuscripts. Hard-copy only submissions cannot be accepted or returned.

The manuscript should:

- show the article's title, authors' names, affiliation, and email addresses (these will be included on the online page), and postal address and telephone numbers (these will be removed from the final online page)
- abstract should not exeed 150 words
- paper should not exceed 5,000 words (20,000 for occasional papers)
- include an abstract of about150 words
- include 5 keywords describing article
- be in final form ready for reviewing
- conform to the APA Reference System

General Principles for layout:

- Papers should be prepared using the template *iejbase.doc* with Microsoft WORD on Mac or PC
- Authors with WORD 98 or later and a basic understanding of web page construction should also save their paper in web format (**.html*)
  incorporating any hypertext features.
- Use end-notes, with one set of numbers running through the whole article.
- Use italics (not underline) for emphasis.
- Do not create additional headers and footers in the template document *iejbase.doc*, just modify appropriately.
- Do not use page numbers in text (web pages do not follow conventional page numbering systems).

If you have any questions regarding layout and presentation, please contact the IEJ Online Editor by email on: onlineeditor@iej.cjb.net

FUIIE

The publishers have arranged for the Flinders University Institute of International Education to produce and host an electronic mirror site containing all material in the Journal, accessed free of charge. There are no restrictions on downloading or printing single copies. The electronic version also provides readers with a search tool.

International Education Journal Vol 1, No 1, 1999
http://iej.cjb.net

Changes in students' mathematics achievement in Australian lower secondary schools over time

Tilahun M. Afrassa
Flinders University of South Australia *tilahun@ssabsa.sa.gov.au*

John P. Keeves
Flinders University of South Australia *john.keeves@flinders.edu.au*

This paper is concerned with the analysis and scaling of mathematics achievement data over time by applying the Rasch model using the QUEST (Adams & Khoo, 1993) computer program. The mathematics achievements of the students are brought to a common scale. This common scale is independent of both the samples of students tested and the samples of items employed. The scale is used to examine the changes in mathematics achievement of students in Australia over 30 years from 1964 to 1994. Conclusions are drawn as to the robustness of the common scale and the changes in students' mathematics achievements over time in Australia.

CHANGES IN MATHEMATICS ACHIEVEMENTS OVER TIME

Over the past four decades researchers have shown considerable interest in the study of student achievement in mathematics at all levels across educational systems and over time. Many important conclusions can be drawn from various research studies about students' achievement in mathematics over time. Willett (1997, 327) argues that by measuring change over time, it is possible to map phenomena at the heart of the educational enterprise. In addition, he argues that education seeks to enhance learning, and to develop change in achievement, attitudes and values. It is Willett's belief that "only by measuring individual change is it possible to document each person's progress and, consequently, to evaluate the effectiveness of educational systems" (Willett, 1997, 327). Therefore, the measurement of changes in achievement over time is one of the most important tools for finding ways and means to improve the educational system of a country.

Since, Australia participated in the 1964, 1978 and 1994 IEA Mathematics Studies, it should be possible to examine the mathematics achievement differences over time across the 30-year time period. The IEA Mathematics Studies were conducted in Australia under the auspices of the International Association for the Evaluation of Educational Achievement (IEA).

Therefore, the purpose of this study is to investigate changes in achievement in mathematics of Australian lower secondary school students between 1964, 1978 and 1994.

In this paper the results of the Rasch analyses of the mathematics achievement of the 1964, 1978 and 1994 Australian students who participated in FIMS, SIMS and TIMS are presented and discussed. The paper is divided into eight sections. The sampling procedures used on the three occasions are presented in the first section, while the second section examines the measurement procedures employed in the study. The third section considers the statistical procedures applied in the calibration and scoring of the mathematics tests. The fourth section assesses whether or not the mathematics items administered in the studies fit the Rasch model. Section five discusses the equating procedures used in the study. The comparisons of the achievement of FIMS, SIMS and TIMS students are presented in the next section. The last section of the paper examines the findings and conclusions drawn from the study.

SAMPLING PROCEDURE

Table 1 shows the Target Populations of the three international studies conducted in Australia. In the First IEA Mathematics Study (FIMS), conducted in 1964, two groups of students participated, 13-year-old students in Years 7, 8 and 9, referred to as FIMSA, and students in Year 8 of schooling referred to as FIMSB. The total number of students taking part was 4320 (see Table 1).

In the first study only government schools in New South Wales (NSW), Victoria (VIC), Queensland (QLD), Western Australia (WA) and Tasmania (TAS) participated. In the Second IEA Mathematics Study (SIMS), which was administered in 1978, nongovernment schools and the Australian Capital Territory (ACT) and South Australia (SA) were also involved as well as those states that participated in FIMS. Thus in SIMS government and nongovernment school students in six states and one territory were involved. The total number of participants was 5120 students (see Table 1).

Meanwhile, in the Third IEA Mathematics Study (TIMS), which was conducted in 1994, government and nongovernment school students in all states and territories including Northern Territory were involved. The total number of students tested was 12,852 (see Table 1).

Table 1. Target populations in FIMS, SIMS and TIMS

Target Population	Label	Size	Sampling Procedure	Primary Unit	Secondary Unit	Design Effect	Effective Sample Size
13-year-old_R	FIMSA	2917	SRS	School	Student	7.66	380
Grade 8_R	FIMSB	3081	SRS	School	Student	11.82	261
Total-G_R	FIMST	4320	SRS	School	Student	11.11	389
13-year-old_R	SIMS	3038	PPS	School	Student	5.40	563
Total-G&NG	SIMST	5120	PPS	School	Student	7.00	731
Grade 8 _R	TIMS	3786	PPS	School	Class	16.52	229
Total-G&NG	TIMST	12852	PPS	School	Class	30.59	420

SRS = Simple-Random-Sample of schools and students within schools; PPS = Probability-proportional -to-size sample; R = Restricted to government schools in five states; G = The participating schools were only government schools; G&NG = Both government and nongovernment schools participated

In 1964 and 1978 the samples were age samples and included students from Years 7, 8 and 9 in all participating states and territory, while in TIMS the samples were grade samples drawn from Years 7 and 8 or Years 8 and 9. In ACT, NSW, VIC and TAS Years 7 and 8 students were selected while in QLD, SA, WA and NT samples were drawn from Years 8 and 9.

Therefore, in order to make meaningful comparisons of mathematics achievement over time by using the 1964, 1978 and 1994 data sets, the following steps were taken.

The 1964 students were divided into two groups 13-year-old students in one group and all Year 8 students including 13-year-old students at that year level as the second group since in addition to an age sample a grade sample had also been drawn. It is important to observe that 13-year-old students in Year 8 were considered as members of both groups. In the first group students were chosen for their age and in the second group for their year level. The 1978 students were chosen as an age sample and included students from both government or nongovernment schools. In order to make meaningful comparisons between the 1978 sample and the 1964 sample, students from nongovernment schools in all participating states and all students from SA and ACT were excluded from the analyses presented in this paper, although Rosier (1980) and Moss (1982) considered both the total student groups and the restricted government school student groups drawn from only five states.

Meanwhile, in TIMS the only common sample for all states and territories was the Year 8 students. In order to make the TIMS samples comparable with the FIMS samples, only Year 8 government school students in the five states that participated in FIMS are considered as the TIMS data set in this study. All nongovernment school students in the five states and all students in SA, ACT and NT are excluded from the analyses presented in this study, although they have been considered in the report by Lokan, *et al.*(1996).

After excluding schools and the states and territories that did not participate in the 1964 study, two sub-populations of students were identified for comparison between occasions. The two groups were 13-year-old students in FIMSA and SIMS: all were 13-year-old students and were distributed across Years 7, 8 and 9 on both occasions. Hence, these two groups of 13-year-old students were considered to be comparable for the examination of achievement over time, between 1964 and 1978. Whereas for the comparison between FIMSB and TIMS the other sub-populations consisted of 1964 and 1994 Year 8 students. Students in both groups were at the same year level, although there were differences in the ages between these groups which were tested on the two occasions. Hence, the comparisons in this study are between 13-year-old students in FIMSA and SIMS on the one hand, and FIMSB and TIMS Year 8 students on the other.

MEASUREMENT PROCEDURES EMPLOYED IN THE STUDY

In this study the procedures employed to measure mathematics achievement level of students on the three occasions involved the use of the Rasch model to scale students' responses to the mathematics test items. The tests included both multiple choice and constructed response items, and in the 1994 testing program both dichotomous and polychotomous scoring procedures were employed for the constructed response items.

Use of Rasch model

The Rasch model has been shown to be the most robust of the item response models (Sontag, 1984), and was used in this study primarily to equate students' performance in mathematics on a common scale.

Unidimensionality

In order to employ the Rasch model for calibrating the items in the mathematics tests it was necessary to examine whether or not the items were unidimensional since the unidimensionality of items is one of the requirements for the use of the Rasch model (Hambleton and Cook, 1977; Anderson, 1994). If the items were found not to satisfy the condition of unidimensionality, it would not be possible to employ the Rasch procedures in the calibration of the tests. Hence, a literature search was undertaken to examine whether or not the test developers had examined the dimensionality of these items, and to the investigators' knowledge the items had at no time been examined for unidimensionality, although Peaker (1969) had considered how the part scores should be weighted.

Consequently, confirmatory factor analysis procedures were employed to test the unidimensionality of the mathematics test items. Confirmatory factor analysis is a statistical procedure employed for investigating relations between a set of observed variables and the underlying latent variables (Byrne, 1989; Kim & Mueller, 1978a, 1978b; Long, 1983; Spearritt, 1994). Thus, confirmatory factor analysis assumes that the observed variables are derived from some underlying source variables (Kim & Mueller, 1978a). Factor analysis may also be used as an appropriate method for determining the minimum number of hypothetical variables that would account for the observed covariation, and thus as a means of exploring the data for possible data reduction (Kim & Mueller, 1978a). However, one of the main purposes of confirmatory factor analysis is to examine and test the common underlying dimensions associated with a number of observed variables.

The results of the confirmatory factor analyses of FIMS and SIMS data sets revealed that a nested model in which the mathematics items were assigned to three specific correlated first-order factors of Arithmetic, Algebra and Geometry as well as a general higher order factor, which was labelled as *Mathematics* provided the best fitting model. In addition, in the confirmatory factor analyses undertaken, no evidence was found to reject the assumption of the existence of one general higher order factor involved in the mathematics tests, in so far as in the nested model the *Mathematics* factor extracted more of the total variance than did the specific lower-order factors taken together. Therefore, the mathematics test items in the FIMS and SIMS studies are considered to satisfy the requirement of unidimensionality. The item cluster-based design procedure (Adams and Gonzalez, 1996) employed in the construction of the TIMS data sets would seem to preclude the use of confirmatory factor analysis to test the unidimensionality of the TIMS data set.

THE STATISTICAL PROCEDURES EMPLOYED IN THE STUDY

In this section the statistical procedures employed in the study are discussed.

Effect size

In this paper both the standardized effect size and the magnitude of effect on the calibrated scales are used to examine the level of practical significance of the differences between FIMS, SIMS and TIMS in mathematics achievements over time. The following formula was employed to calculate an effect size value.

$$\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2 + s_2^2}{2}}}$$

Where $\bar{x}_1$ = estimated mean score for group one;

$\bar{x}_2$ = estimated mean score for group two;

s_1 = standard deviation of the mean of group one; and

s_2 = standard deviation of the mean of group two.

In this study effect size values less than 0.20 are considered as trivial, while values between 0.20 and 0.50 are considered as small. Furthermore, effect size values between 0.50 and 0.80 are taken as medium and values above 0.80 are treated as large (Cohen, 1991; Keeves, 1992).

Growth between grade levels

It is possible since the TIMS project tested at two adjacent grades to estimate the gain between the lower grade and the upper grade for the Australian sample and thus to interpret the calibrated effect size in terms of a year of mathematics learning at the lower secondary school level. Afrassa (1998) has reported that the growth in achievement per year in mathematics achievement in Australian lower secondary schools was 37 centilogits. This value is equivalent to an effect size of 0.30. Keeves (1992) has indicated that an effect size of 0.30 was also found to be equivalent to a year's learning in science at the lower secondary school level. Therefore, this additional information allows the differences between the achievement level of the different groups to be interpreted in terms of practical significance rather than depending solely on statistical significance.

The t-test

In order to determine the level of statistical significance between the mean scores on FIMS, SIMS and TIMS in mathematics achievement a *t* statistic was calculated, which took into account errors from three scores: (a) sampling errors; (b) error of calibration, and (c) equating error. Further comment on the estimation of these errors is given in Appendix A.

Treatment of omits and non-responses

Issues regarding the occurrence and handling of missing data in achievement tests of the kind employed in these three mathematics studies were considered. However, the results of the Rasch analyses did not show marked differences between ignoring the missing data or treating the missing data as wrong during the calibration and scoring. Therefore, for both calibration and scoring purposes it was decided to treat the missing data as wrong. While all the items in the mathematics tests were employed for scoring purposes, for calibration purposes only those items that fitted the Rasch scale were considered. The main justification for the use of these procedures would seem to lie in the greater number of misfitting items when the procedure that involved the ignoring of the missing data was tested with the SIMS data. Consequently, the procedure that involved treating missing data as wrong was chosen for this study.

Treatment of zero and perfect scores

The QUEST computer program (Adams and Khoo, 1993) by default does not process cases with perfect and zero scores, because both groups do not provide information for the calibration of the scale. Cases with perfect scores are those cases who provided correct responses for all the items, while cases with zero scores are those cases who provided wrong responses for all items. Hence, in order to include those cases with perfect and zero scores in the calculation of the mean and standard deviation of the mathematics achievement test scores for each student sample, the values of perfect and zero scores were calculated by extrapolation from the logit table produced by the QUEST computer program (see Appendix B). Subsequently, the SPSS 6.1. (Norusis & SPSS Inc, 1990) computer program was used to calculate the case estimate mean scores and standard deviations with appropriate weights, and the WesVarPC 2.11 (Brick, Broene, James and Severynse, 1997) computer program was employed for calculating the standard error of the mean values, again with appropriate weighting of the data, and with allowance made for the fact that all samples were of a stratified cluster sample design.

Developing a common mathematics scale

The calibration of the mathematics data permitted a scale to be constructed that extended across the three groups, namely FIMS, SIMS and TIMS students on the mathematics scale. The fixed point of the scale was set at 500 with one logit, the natural metric of the scale, being set at 100 units. The fixed point of the scale, namely 500 was taken as the mean of the difficulty level of the calibrated items in the FIMS test administered in 1964. The mathematics scale constructed in this way for all different sample groups of students in FIMS, SIMS and TIMS is presented in Figures 1, 2 and 3, with 100 scale units (centilogits) being equivalent to one logit.

Conclusion

In the last two sections the scaling and the statistical procedures employed in the study are discussed. The Rasch model was the major scaling procedure employed. The effect size and the *t*-test were employed for comparing the mean values of different groups of students. With respect to the missing data a decision was made, from a study of the results of the Rasch analyses, for both calibration and scoring purposes to treat the missing data as wrong, while for calibration purposes only those items that fitted the Rasch scale were employed.

RASCH ANALYSIS

Three groups of students namely FIMS (4320), SIMS (5120) and TIMS (7392) were employed in the calibration and scoring analyses. The necessary requirement for calibration in Rasch scaling is that the items and persons must fit the Rasch scale. Items and persons which do not fit the scale must be deleted in calibration. In order to examine whether or not the items and persons fitted the scale, it was also important to evaluate both the item fit statistics and the person fit statistics. The results of these analyses are presented below.

Item fit statistics

One of the key item fit statistics is the infit mean square (INFIT MNSQ). The infit mean square measures the consistency of fit of the students to the item characteristic curve for each item with weighted consideration given to those persons close to the 0.5 probability level. The acceptable range of the infit mean square statistic for each item in this study was taken to be from 0.77 to 1.30 (Adams and Khoo, 1993). Values outside this acceptable range, that is above 1.30 indicate that these items do not discriminate well, and below 0.77 the items provide redundant information. Hence, consideration must be given to excluding those items that are outside this range. In calibration, items that do not fit the Rasch model and which are outside of the acceptable range must be deleted from the analysis (Rentz and Bashaw, 1975; Wright and Stone, 1979: Kolen and Whitney, 1981; Smith and Kramer, 1992). Hence, in FIMS two items (Items 13 and 29), in SIMS two items (Items 21 and 29) and in TIMS one item [(Item T1b No 148) with one item (no 94) having been excluded from the international TIMSS analysis] were removed from the calibration analyses due to the misfitting of these items to the Rasch model.

Case estimates

The other way of investigating the fit of the Rasch scale to data is to examine the estimates for each case. The case estimates give the performance level of each student on the total scale. In order to identify whether the cases fit the scale or not, it is important to examine the case OUTFIT mean square statistic (OUTFIT MNSQ) which measures the consistency of the fit of the persons to the student characteristic curve for each student, with special consideration given to extreme items. In this study, the general guideline used for interpreting **t** as a sign of misfit is if t> 5 (Wright and Stone, 1979, 169). Thus, if the OUTFIT MNSQ value for a person has a $|t - value|$ greater than 5, that person does not fit the scale and is deleted from the analysis. In this analysis no person was deleted, because the $|t - value|$ for all cases was less than 5. However, students with a zero score or with a perfect score were automatically excluded from the calibration procedure.

Conclusion

In summary, the results of the infit mean square indices, revealed that 68 out of 70 items for FIMS, 70 out of 72 items for SIMS and 156 out of 157 items for TIMS data sets fitted the Rasch model. In addition, the evidence indicated that for all cases, the responses of the students sampled fitted the Rasch model, except for those students who had perfect or zero scores.

EQUATING OF MATHEMATICS ACHIEVEMENT OVER TIME

Equating of the mathematics tests requires common items between occasions, that is between FIMS, SIMS and TIMS. Wright and Stone (1979) have recommended that 10 to 20 (17 to 34 per cent of the items in each test) items should be employed for equating two different test forms consisting of 60 items each. Meanwhile, Hambleton *et al.*, (1991) suggested approximately between 20 and 25 per cent of the number of the items in the tests should be common. However, Smith and Kramer (1992) have argued that as few as a single item is required.

In this study, the number of common items in the mathematics test for FIMS and SIMS data sets was 65. For the mathematics tests the common items formed approximately 93 per cent of the items for FIMS, and 90 per cent for SIMS. Thus, the common items in the mathematics test for these two occasions were well above the percentage ranges proposed by Wright and Stone (1979) and Hambleton *et al.* (1991).

There were also some items which were common for FIMS, SIMS and TIMS data sets. Garden and Orpwood (1996, 2-2) reported that achievement in TIMS was intended to be linked with the results of the two earlier IEA studies. Thus, in the TIMS data set there were nine items which were common for the three occasions. Therefore, it was possible to claim that there were sufficient numbers of common items to equate the mathematics tests on the three occasions.

Rasch model equating procedures were employed for equating the three data sets. Rentz and Bashaw (1975), Beard and Pettie (1979), Sontag (1984) and Wright (1995) have argued that Rasch model equating procedures are better than other procedures for equating achievement tests. All three types of Rasch model equating procedures, namely concurrent equating, anchor item equating and common item difference equating were used for equating the three data sets.

Concurrent equating was employed for equating the data sets from FIMS and SIMS. In this method, the 65 common items between FIMS and SIMS were combined into one data set. Hence, the analysis was done on a single data file. Only one misfitting item was deleted at a time so as to avoid dropping some items that might eventually prove to be good fitting items. The acceptable infit mean square values were between 0.77 and 1.30 (Adams and Khoo, 1993). The concurrent equating analyses revealed that among the 65 common items 64 items fitted the Rasch model. Therefore, the threshold values of these 64 items were used as anchor values in the anchor item equating procedures employed in the scoring of the FIMS and SIMS data sets separately. Among the 64 common items, nine were common to the FIMS, SIMS and TIMS data sets. The threshold values of these nine items generated in this analysis are presented in Table 2 and were used in equating the FIMS data set with the TIMS data sets.

The design of TIMS was different from FIMS and SIMS in two ways. In the first place, only one mathematics test was administered in both FIMS and SIMS, however, in the 1994 study the test included mathematics and science items and the study was named TIMSS (Third International Mathematics and Science Study). The other difference was that in the first two international studies, the test was designed as one booklet. Every participant used the same test booklet. Whereas in TIMSS, a rotated test design was used. The test was designed in eight booklets.

Garden and Orpwood (1996, 2-16) explained the arrangement of the test in eight booklets as follows.

This design called for items to be grouped into "clusters", which were distributed (or "rotated") through the test booklets so as to obtain eight booklets of approximately equal difficulty and equivalent content coverage. Some items (the core cluster) appeared in all booklets, some (the focus cluster) in three or four booklets, some (the free-response clusters) in two booklets, and the remainder (the breadth clusters) in one booklet only. In addition, each booklet was designed to contain approximately equal numbers of mathematics and science items.

All in all there were 286 unique items that were distributed across eight booklets for Population 2 (Adams and Gonzalez, 1996, 3-2).

In order to investigate the level of mathematics achievement in TIMS, it is necessary to find ways and means for equating these eight booklets. Furthermore, in order to employ any kind of test equating procedure there must be common items between the different booklets. Garden and Orpwood (1996) reported that the core cluster items (six items for mathematics) were common to all booklets. In addition, the focus cluster and free-response clusters were common to some booklets. Thus, it was possible to equate these eight booklets and report the achievement level in TIMS on a common scale.

Hence, from among the Rasch model test equating procedures, concurrent equating was chosen for equating these eight booklets. The purposes of the test equating in TIMS was to investigate the mathematics achievement level of Australian students in TIMS and subsequently to compare the result with FIMS and SIMS data sets.

Consequently, concurrent equating procedures were employed for the TIMS data set. The result of the Rasch analysis indicated that only one item was deleted from the analysis. The item which was deleted from the analysis was Item T1b (No 148) which was below the critical value of 0.77. All other items fitted well the Rasch model. Hence, out of 157 items, 156 of the TIMS test items fitted well the Rasch model. From the output of the concurrent equating, it was possible to obtain the threshold values of the nine common items in TIMS. These threshold values are shown in Table 2.

The next step involved the equating of the FIMS data set with the TIMS data set using the common item difference equating procedure. In this method the threshold values of the FIMS test generated by the QUEST computer program (Adams and Khoo, 1993) for each state are first subtracted from threshold values of the TIMS test. Then the differences are summed and divided by the number of anchor test items to obtain a mean difference between FIMS and TIMS for each state (see Table 2). The interesting point to be mentioned is that the difference in threshold values between the two occasions in each of the five states is generally similar. The difference between the state with the highest mean threshold difference and the lowest mean score difference was only 0.18. The highest mean threshold estimate was registered in WA (1.14) while the lowest was in NSW and VIC, the mean difference score for both states was 0.96 (see Table 2). This result revealed that the common items in the two tests behaved similarly in all the five states.

The grand mean difference was calculated by adding the five states mean difference threshold estimates and dividing them by five. The resulting mean difference across states was 1.03. The grand mean of the differences (1.03) is called the equating constant. The equating constant is subsequently employed in the calculation of the TIMS scores on the FIMS scale. That is the equating constant was subtracted from the Rasch estimated mean score on the TIMS for each state to obtain the adjusted mean value of TIMS for each state. The comparisons of achievement over time in the five Australian states using the weighted Rasch estimated scores of 1964, 1978 and 1994 for each state are discussed in the next section.

Table 2. Descriptive statistics of the common item difference equating procedure employed in FIMS and TIMS

NSW			**VIC**			**QLD**			**WA**			**TAS**		
F	T	T - F	F	T	T - F	F	T	T - F	F	T	T - F	F	T	T - F
0.35	1.00	0.65	0.13	0.45	0.32	0.62	1.17	0.55	0.15	1.29	1.14	0.00	0.87	0.87
1.10	1.86	0.76	0.06	1.96	1.90	-0.61	0.96	1.57	0.46	1.90	1.44	0.27	2.09	1.82
-2.90	-0.73	2.17	-2.71	-0.79	1.92	-2.64	-0.79	1.85	-2.82	-0.76	2.06	-2.85	-0.87	1.98
-0.44	1.77	2.21	0.51	1.30	0.79	0.00	1.60	1.60	0.40	1.39	0.99	0.22	1.26	1.04
-1.41	-0.49	0.92	-1.24	-0.48	0.76	-1.58	-0.52	1.06	-0.81	-0.30	0.51	-1.28	-0.10	1.18
-1.12	-0.92	0.20	-1.01	-0.83	0.18	-0.63	-0.81	-0.18	-1.11	-0.81	0.30	-0.94	-1.10	-0.16
0.94	0.68	-0.26	0.31	1.09	0.78	1.49	2.24	0.75	0.43	1.76	1.33	0.85	1.53	0.68
-0.05	1.54	1.59	0.11	1.64	1.53	-0.54	1.64	2.18	-0.08	1.73	1.81	-0.07	1.56	1.63
-0.03	0.39	0.42	0.13	0.58	0.45	0.09	0.31	0.22	-0.04	0.61	0.65	-0.04	0.23	0.27
Mean		8.66			8.63			9.60			10.23			9.31
		0.96			0.96			1.07			1.14			1.03
Grand Mean = 1.03				**SD = 0.699**			**SE = 0.104**							

The common items were numbered differently in FIMS (F) and TIMS (T), in FIMS the numbers were 12, 26, 31, 32, 33, 36, 38, 54 and 67, while in TIMS the number of the common items were 70, 62, 6, 136, 129, 92, 42, 31 and 39 respectively. TIMS - FIMS (T-F) shows the difference between threshold value of an item in the TIMS and FIMS mathematics tests.

Comparisons of achievement over time

The comparisons of the performance of students on the mathematics test for the three occasions were undertaken for two different subgroups namely: (a) 13-year-old students in government schools, who participated in the FIMS and SIMS studies; and (b) Year 8 government school students who participated in the FIMS and TIMS studies. All SIMS students were 13-year-old students. Meanwhile, some of the FIMS students were 13-year-old students, while others were younger and/or older students who were in Year 8. Therefore, for comparison purposes the FIMS students were divided into two groups, namely: (a) FIMSA - involving all 13-year-old students, and (b) FIMSB - including all Year 8 students. Thus, FIMSA students' results could be compared with SIMS students in the government schools of five states, because all students were 13-year-olds. In the TIMS analyses a decision was made to include only Year 8 students, because they were the only group of students who were common to all participating states. Thus TIMS Year 8 government school students from the five states involved could be compared with FIMSB students, because in both groups the students were at the same year level.

COMPARISON BETWEEN STUDENTS IN MATHEMATICS ACHIEVEMENT OVER TIME

The first part of this section addresses the comparisons between FIMSA and SIMS, while the second part discusses the comparison between FIMSB and TIMS.

Comparison between 13-year-old students' mathematics achievement over time

In this section the achievement of 13-year-old Australian students who participated in FIMS and SIMS are compared. Table 3 presents the results of the analyses of the comparison between the two occasions. The first and second panels of the table show the participating states, the estimated case means of the 13-year-old students, the standard deviations and the standard error values, the sample sizes, design effects and effective sample sizes for FIMS and SIMS respectively. While the third panel presents the estimated mean differences between the two groups, the effect sizes and t-values of the differences and the significance levels.

Table 3. Comparisons between FIMS and SIMS 13-year-old Government School Students

13-year-old students	**FIMS**						
State	Mean score	SD	SE	Sample Size	DEFF	ESS	Age
State A	458	99.3	9.5	640	5.84	110	13.4
State B	483	101.5	13.4	723	12.50	58	13.3
State C	423	89.8	11.9	402	7.13	56	13.5
State D	459	90.2	6.7	678	3.72	182	13.5
State E	444	85.6	8.5	474	4.64	102	13.5
Australia	460	96.2	4.9	2917	7.66	381	13.4
13-year-old students	**SIMS**						
State	Mean score	SD	SE	Sample Size	DEFF	ESS	Age
State A	442	102.4	6.4	675	2.66	254	13.4
State B	472	105.8	8.4	643	4.09	157	13.4
State C	428	94.0	7.6	529	3.53	150	13.5
State D	423	98.3	9.4	618	5.63	110	13.5
State E	444	95.3	6.2	573	2.42	236	13.4
Australia	441	102.2	4.3	3038	5.40	563	13.4

State	**SIMS-FIMS**	**Effect Size**	**t-value**	**Significance Level**
State A	-16	-0.16	-1.39	NS
State B	-11	-0.11	-0.70	NS
State C	05	0.05	0.14	NS
State D	-36	-0.37	-3.12	0.01
State E	00	0.00	0.00	NS
Australia	-19	-0.19	-2.91	0.01

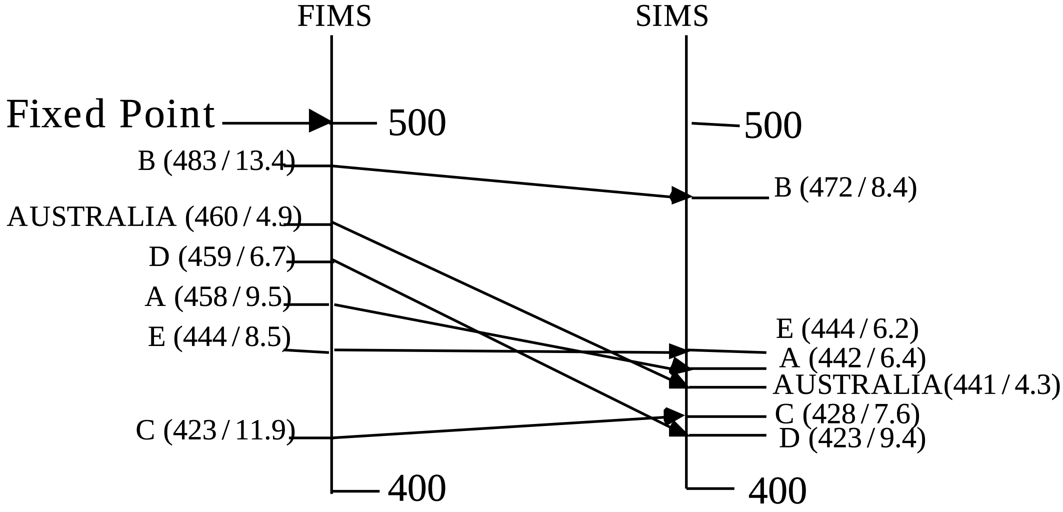
SD = Standard deviation; SE = Standard error; ESS = Effective sample size; DEFF = Design effect

State A

When the 1964, 13-year-old State A students estimated mean score is compared with the 1978 same age group students in the same state, the mean score of the 1964 students (458) was higher than that of their 1978 peers (442). The difference was 16 centilogits (see Figure 1 and Table 3). The differences in standard deviation and standard error values for the two groups were slight, while the design effect was larger in 1964 than in 1978. The effect size was trivial (0.16) and the t-value was 1.39. The estimated mean difference indicated that the mathematics achievement of 13-year-old students in State A had declined over time. However, the effect size and t-values showed that the difference was not practically or statistically significant. Hence, it is not possible to conclude that there was a significant decline in mathematics achievement in State A at the 13-year-old student level.

State B

Table 3 and Figure 1 indicate that when the estimated mean score of the 13-year-old State B students, who participated in the 1964 First IEA Mathematics Study, is compared with the mean score of the 1978 same age group students who participated in the Second IEA Mathematics Study, the 1964 students (483) were found to be higher achievers than their 1978 peers (472). However, the difference was slight, 11 centilogits (see Figure 1 and Table 3). The difference in the standard deviation values for the two groups was also slight, while the standard error and design effect were large. Both were larger in 1964 than in 1978. The effect size was trivial (0.11) and the t-value was 0.70. The estimated mean difference in scores indicated that the mathematics achievement of 13-year-old students in State B declined only slightly over time, since, the effect size and t-value showed that the difference was not practically or statistically significant. Therefore, there was no significant decline in mathematics achievement in State B at the 13-year-old student level.



Fixed point = Mean difficulty level of FIMS test
100 units= 1 logit; 1 unit = 1 centilogit
Values in brackets are Rasch estimated mean scores and standard errors of the mean respectively with State A (A), State B (B), State C (C), State D (D) and State E (E)

Figure 1. Comparison of Achievement in Mathematics between 1964 and 1978 in Australia of 13-year-old students

State C

The next state that was considered in the comparison between 1964 and 1978 was State C. The estimated mean score of the 1964 13-year-old State C students was 423, meanwhile, the same age group students in 1978 scored 428 (see Table 3 and Figure 1). The mean score difference between the two groups was five centilogits in favour of the 1978 students. This indicated that unlike State A and State B, in State C the achievement level of 13-year-old students increased over time (see Figure 1 and Table 3). The difference in standard deviation values for the two groups was slight, while the standard errors and the design effects were larger in 1964 than in 1978. The effect size was trivial (0.05) and the t-value was 0.14. The estimated mean difference indicated that the mathematics achievement of 13-year-old students in State C had improved very slightly over time. However, the effect size and t-value showed that the difference was neither practically nor statistically significant. Hence, it was possible to conclude that there was no significant improvement in mathematics achievement in State C at the 13-year-old student level between 1964 and 1978.

State D

State D was one of the five Australian states that participated in both the 1964 and 1978 IEA mathematics studies. The estimated mean score value of the 1964 13-year-old State D students was compared with the 1978 same age group government school students in that state. The mean score difference between students in the two studies was 36 centilogits and the difference was in favour of the 1964 students (see Figure 1 and Table 3). This showed that the achievement of the 1978 students was noticeably lower than that of the 1964 students. In other words, achievement had declined from 1964 to 1978 in the State D government schools. The differences in standard deviation and standard error values for the two groups were slight, while the design effect was larger in 1978 than in 1964. The effect size was small (0.37) and the t-value was 3.12. The estimated mean difference indicated that the mathematics achievement of 13-year-old students in State D government schools had declined over time. In addition, the effect size and t-values also showed that the difference was both practically and statistically significant at the 0.01 level. Hence, it would seem possible to conclude that there was a significant decline in mathematics achievement in State D in government schools at the 13-year-old student level from 1964 to 1978, and that the decline in mathematics achievement represented more than one year's learning of mathematics in the lower secondary schools of Australia.

State E

State E was the last state for the comparison of performance between the 1964 and 1978 13-year-old students who participated in FIMS and SIMS respectively. When the estimated mean score value of the 1964, 13-year-old State E students was compared with the 1978 same age group government school students in the same state there was no difference in their mean scores. The mean scores of both groups was 444 (see Figure 1 and Table 3). There was no difference between the achievement of 13-year-old students in State E government schools between 1964 and 1978. The differences in standard deviation and standard error values for the two groups were slight, while the design effect was larger in 1964 than in 1978. The effect size and the t-value were both 0.00. Hence, it would seem possible to conclude that there was no difference in mathematics achievement in State E government schools at the 13-year-old students level over the 14-year period.

The above comparisons were for 13-year-old students in the five states between 1964 and 1978. Among the five states, even if it was not statistically significant, it was only in State C, that achievement over time improved slightly. However, there was no difference between the two occasions in State E. Moreover, the remaining three states, that is in State A, State B and State D, achievement over time declined, but the decline was significant at the 0.01 level only for State D.

Australia

The results addressed above, led to the comparison of the overall Australian 13-year-old students between the two occasions. The estimated mean score difference between the two occasions was 19 centilogits and the difference was in favour of the 1964 13-year-old Australian students. This revealed that the mathematics achievement of Australian students declined from 1964 to 1978. The differences in standard deviation and standard error values for the two groups were small, while the design effect was slightly larger in 1964 than in 1978. The effect size was not inconsiderable (0.19) and the t-value was 2.91. Hence, the mean difference was statistically significant at the 0.01 level (see Table 2 and Figures 1 and 3). Moreover, in Australia the mathematics achievement level of the 13-year-old students declined over time, between 1964 and 1978, to an extent that represented approximately half (19/37) of a year of mathematics learning.

In conclusion, the comparisons of the mathematics achievement of the 13-year-old students between 1964 and 1978 in the five Australian states and overall in Australia revealed that in three states and in Australia overall achievement had declined over time. Statistically significant declines were recorded only for State D and for Australia overall. There was no difference between the two occasion for State E students. While, an improvement was recorded for State C, the increase was slight, and it was not statistically significant. The next section presents the comparison of mathematics achievement at the Year 8 level between 1964 and 1994 for students in the government schools of the five states.

Comparison between Year 8 students' mathematics achievement over time

In this section the achievement levels of Year 8 students between 1964 and 1994 are compared. The results of the comparisons of students by state are presented in Table 4 and Figure 2.

State A

The first state which was selected for comparison was State A. The estimated mean score difference between the two occasions at the Year 8 level in State A was two centilogits, the difference was in favour of the 1964 students (see Figure 2 and Table 4). This result indicated that the mathematics achievement at the Year 8 level had declined very slightly between 1964 and 1994 in State A government schools. The effect sizes and t-values were too small to be considered, and this decline in achievement over time in State A schools at the Year 8 level was not found to be statistically significant.

State B

State B was the next state that participated in the three international mathematics studies. When the estimated mean scores of the FIMSB and TIMS groups were compared, the 1964 students mean score was noticeably higher than that of the 1994 students. This revealed that mathematics achievement over time had declined in State B schools at the Year 8 level. The standard deviation, standard error and the design effect were larger in 1994 than in 1964. The effect size (0.83) and t-value (4.22) were large. Hence, the decline in mathematics achievement at the Year 8 level between 1964 and 1994 was practically and statistically significant at the 0.01 level. Moreover, it should be noted that the decline in mathematics achievement in this state represented more than two years of learning (82/37=2.2).

State C

The mathematics achievement difference between 1964 and 1994 in State C schools at the Year 8 level was small. The mean difference was 27 centilogits in favour of the 1964 Year 8 students. The result indicated that the Year 8 students' mathematics achievement in State C had declined over time. The standard deviation, standard error and the design effect were larger in 1994 than in 1964. The effect size (0.28) was small, but the t-value (1.46) was not significant. Thus, the t-value indicated that the decline in mathematics achievement between the 1964 and 1994 Year 8 State C students was not statistically significant. Hence, it would seem possible to conclude that there was no statistically significant decline in achievement between 1964 and 1994 in State C at the Year 8 level. However, a substantial decline would seem to have occurred since the difference (0.27) represented approximately three quarters of a year of mathematics learning.

Table 4. Comparisons between FIMS and TIMS Year 8 Government School Student

Year 8	**FIMS**						
State	Mean score	SD	SE	Sample Size	DEFF	ESS	Age
State A	435	78.5	8.1	686	7.37	93	13.0
State B	485	71.2	4.5	731	2.93	249	13.2
State C	429	72.3	6.4	417	2.35	177	13.9
State D	463	87.0	9.3	742	8.51	87	13.7
State E	438	72.3	2.3	505	0.53	953	13.2
Australia	451	82.1	5.1	3081	11.82	261	13.4
Year 8	**TIMS**						
State	Mean score	SD	SE	Sample Size	DEFF	ESS	Age
State A	433	136.0	17.9	1037	17.97	58	14.1
State B	403	119.9	15.7	752	12.90	58	13.6
State C	402	112.3	13.8	411	6.21	66	14.0
State D	402	116.4	10.6	799	6.62	121	14.0
State E	451	112.3	13.8	787	11.82	67	13.5
Australia	420	125.9	8.3	3786	16.52	229	13.9

State	**TIMS-FIMS**	**Effect Size**	**t-value**	**Significance Level**
State A	-2	-0.02	-0.09	NS
State B	-82	-0.83	-4.22	0.01
State C	-27	-0.28	-1.46	NS
State D	-61	-0.59	-3.47	0.01
State E	13	0.14	0.74	NS
Australia	-31	-0.29	-2.16	0.05

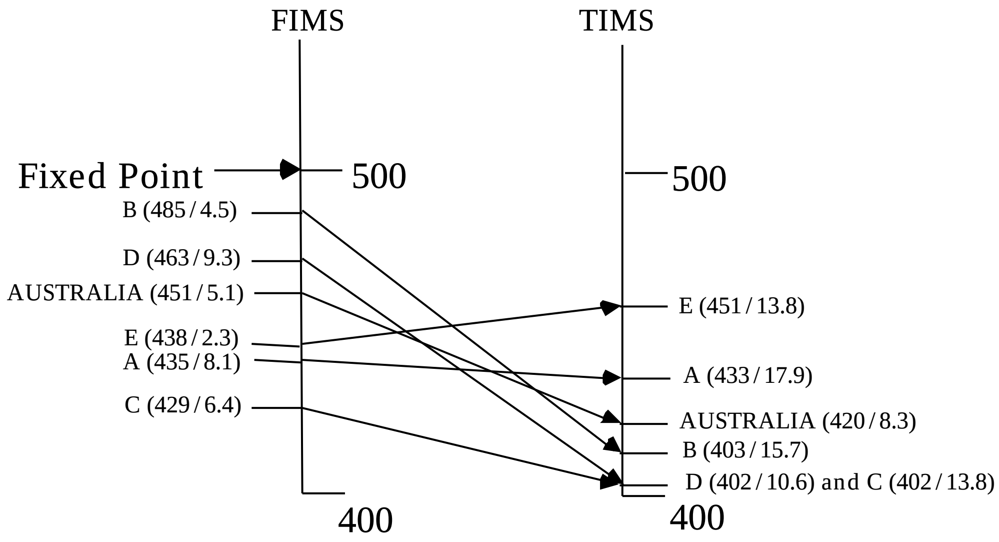
SD = Standard deviation; SE = Standard error; DEFF = Design effect; ESS = Effective sample size

State D

The next comparison was between the State D students, and the mean score difference between 1964 and 1994 Year 8 students was 61 centilogits. The difference was in favour of the 1964 students. This indicated that the mathematics achievement level of Year 8 State D school students had declined by more than a year and a half (61/37) of mathematics learning over the last 30 years. This difference was marked. The effect size was medium (0.59) and the t-value was also large (3.47). Hence, the difference was statistically significant at the 0.01 level. The standard deviation and standard error were larger in 1994 than in 1964. However, the design effect was larger in 1964 than in 1994. Thus, in State D government schools the mathematics achievement level of Year 8 students had declined substantially over the last three decades.

State E

The last state for comparison between the 1964 and 1994 Year 8 students who participated in FIMS and TIMS respectively was State E. When the estimated mean score value of the 1964 State E Year 8 students was compared with that of the 1994 students, the mean score of the 1994 students was higher than that of the 1964 students (see Figure 2 and Table 3). The difference was 13 centilogits. The finding indicated that in State E schools the mathematics achievement level of Year 8 students had improved over the last three decades. The standard deviation, standard error value and design effect were markedly larger in 1994 than in 1964. The effect size (0.14) and the t-values (0.74) were too small to be considered significant. Hence, it would seem possible to conclude that while there was no statistically significant difference in mathematics achievement in State E schools at the Year 8 level between 1964 and 1994. However, some signs of improvement had occurred in marked contrast to the other four states, and that the gain was estimated to be approximately one third of a year (13/37) mathematics learning.



Fixed point = Mean difficulty level of FIMS test; 100 units = 1 logit; 1 unit = 1 centilogit; values in parentheses are Rasch estimated mean scores and standard errors of the mean respectively, with State A (A), State B (B), State C (C), State D (D), and State E (E)

Figure 2. Comparison of Achievement in Mathematics between 1964 and 1994 in Australia

The comparisons in the mathematics achievement level of Year 8 students between 1964 and 1994 in State A, State B, State C, State D and State E revealed that only State E showed improvement in mathematics achievement over the last 30 years. However, the improvement was not found to be statistically significant. Moreover in State B, State C and State D the achievement of Year 8 students had declined over the past three decades. A significant decline was recorded in both State B and State D, although the decline in State C was not statistically significant. The next comparison is between Australian Year 8 students on the two occasions

Australia

The estimated mean score of the 1964 Australian Year 8 students was 451, while, it was 420 in 1994. The difference was 31 centilogits in favour of the 1964 students (see Table 4, Figures 2 and 3). This difference revealed that the mathematics achievement level of Australian Year 8 students had declined over the 30 year period. The standard deviation, standard error and the design effect were larger in 1994 than in 1964. The effect size was 0.29 and the t-value was 2.16. The effect size and t-value indicated that the decline in mathematics achievement between the 1964 and 1994 Year 8 Australian students was marginally significant at the 0.05 level, and the size of the decline was a little less (31/37) than a year of mathematics learning.

COMPARISON OF STANDARD DEVIATION

Table 5 shows the standard deviation values for each state in TIMS and FIMS. There would appear to be a large increase in the spread of scores as measured on the scale of mathematics achievement between 1964 and 1994. This increase may be a consequence of differences in accurate of measurement since in 1964 students answered 70 test items while in 1994 the students answered between 33 and 41 items. However, it would seem more probable that there was greater variability in students' mathematical achievement in 1994 compared with 1964 at the Year 8 level as a consequence of changed teaching and learning practices. This issue warrants further investigation.

COMPARISON OF AGES

Table 4 records the ages of the Year 8 samples of students. It is clear that the decline in student performance can not be attributed to the younger age of the students, since in all samples, the students were older in 1994 at the Year 8 level than in 1964. The increase in age in State A between 1964 and 1994 was the result of a restructuring of the school system that was occurring at about the time of testing in that state in 1964.

Table 5. Comparisons of standard deviation values between FIMS and TIMS

State	TIMS	FIMS	F ratio	Df_T / df_F	p-value
A	136.0	78.5	3.00	58 / 93	<0.01
B	119.9	71.2	2.84	58 / 249	<0.01
C	112.3	72.3	2.41	66 / 177	<0.01
D	116.4	87.0	1.79	121 / 87	<0.05
E	112.3	72.3	2.41	67 / 953	<0.01
Australia	125.9	82.1	2.35	229 /261	<0.01

CONCLUSION

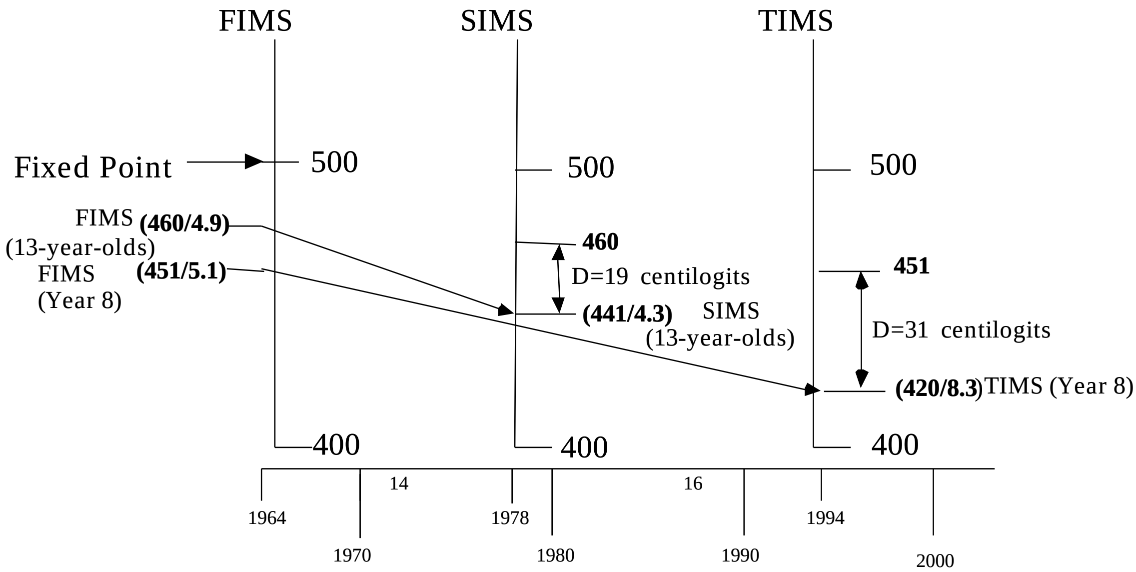
In order to investigate the mathematics achievement level of lower secondary school Australian students over time, three different data sets, namely from the FIMS, SIMS and TIMS studies were analysed. From the three data sets two groups of students were compared. The first comparison was between 13-year-old government school students in five states who participated in FIMS and SIMS. The result of the comparison revealed that only State C showed improvement in mathematics achievement over the 14-year period. However, the improvement was not statistically significant. Furthermore, no achievement difference was found in State E between 1964 and 1978. Meanwhile mathematics achievement showed a decline in State A, State B and State D. Among the three states a significant decline was found only in State D. When the overall Australian students' performance was compared between 1964 and 1978, the mathematics achievement level of the 13-year-old students declined over the 14-year period (see Figure 3).

The second comparison was the mathematics achievement level of Year 8 government school students between 1964 and 1994 in the five states. The findings indicated that State E has improved in mathematics achievement over the last 30 years, however, the improvement was not found to be statistically significant. Whereas, in State A, State B, State C and State D the achievement of Year 8 students had declined over the last three decades. Significant declines were recorded in State B and State D. However, the declines in State A and State C were not significant. When Australian Year 8 government school students who participated in FIMS and TIMS were compared the decline in mathematics achievement level was found to be marginally significant over the last 30-year period (see Figure 3), but of the order of a little less than a year of mathematics learning in Australian lower secondary schools. The findings in both comparisons revealed that the mathematics achievement level of Australian students at the lower secondary school level have declined over the last three decades (see Figure 3). The findings also indicate that there is a need to investigate differences in conditions of learning. Carroll's (1963) model of school learning has guided IEA studies and could guide this investigation. Carroll (1963) has identified five factors that influence school learning. These factors are divided into two levels, namely student and school level factors. The student level factors in Carroll's model are aptitude (home background), ability and perseverance (motivation, attitudes). While the school level factors are time for learning (including homework time for mathematics) and quality of instruction. The investigation demands the use of both:

(1) multivariate analysis, and

(2) multilevel analysis.

Thus, it is necessary to conduct further research to identify the reasons and to recommend solutions for the problems.



Fixed point = Mean difficulty level of FIMS test; 100 units = 1 logit; 1 unit = 1 centilogit
D difference in mathematics achievement between occasions
Values in parentheses are Rasch estimated mean scores and standard errors of the means respectively
$D_{1978-1964}$ = 19 centilogits or _ year of mathematics learning
$D_{1994-1964}$ = 31 centilogits or 4/5 year of mathematics learning

Figure 3. Comparison of Achievement in Mathematics between 1964, 1978 and 1994 in Australia

Acknowledgment

The first author was sponsored by The Flinders University of South Australia Overseas Postgraduate Research Scholarship and the Flinders University Research Scholarship.

REFERENCES

Adams, R. J. & Khoo, S.T. (1993). *Quest- The Iinteractive Test Analysis System.* Hawthorn, Victoria: ACER.

Adams, R. J. & Gonzalez, E. J. (1996). The TIMSS test design. In M.O. Martin & D.L. Kelly (eds), *Third International Mathematics and Science Study Technical Report vol. 1*, Boston: IEA, pp. 3-1 - 3-26.

Afrassa, T. M. (1998). *Mathematics Achievement at the Lower Secondary School Stage in Australia and Ethiopia: A Comparative Study of Standards of Achievement and Student Level Factors Influencing Achievement.* Unpublished PhD thesis, The Flinders University of South Australia, Adelaide, Australia.

Anderson, L. W. (1994). Attitude Measures. In T. Husén (ed), *The International Encyclopedia of Education, vol. 1, (second ed.)*, Oxford: Pergamon, pp. 380-390.

Beard, J. G. & Pettie, A. L. (1979). A comparison of Linear and Rasch Equating results for basic skills assessment Tests. Florida State University, Florida: ERIC.

Brick, J. M., Broene, P., James, P. & Severynse, J. (1997). *A user's guide to WesVarPC. (Version 2.11).* Boulevard, MD: Westat, Inc.

Byrne, B. M. (1989). *A Primer of LISLEL Basic Applications and Programming for Confirmatory Factor Analytic Models.* New York: Springer-Verlag.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112 (1), 155-159.

Garden , R. A. & Orpwood, G. (1996). Development of the TIMSS achievement tests. In M 0. Martin & D L Kelly (eds), *Third International Mathematics and Science Study Technical Report Volume 1: Design and Development*, Boston: IEA, pp. 2-1 to 2-19.

Hambleton, R. K.& Cook, L. L. (1977). Latent trait models and their use in the analysis of educational test data. *Journal of educational measurement*, 14 (2), 75-96.

Hambleton, R. K., Zaal, J. N.& Pieters, J. P. M. (1991). Computerized adaptive testing: theory, applications, and standards. In R.K Hambleton & J.N. Zaal (eds), *Advances in Educational and Psychological Testing*, Boston, Mass.: Kluwer Academic Publishers, pp. 341-366.

Keeves, J. P. (1992). The design and conduct of the second science study. In J.P. Keeves (ed.), *The IEA Study of Science III: Changes in Science Education and Achievement: 1970 to 1984*, Oxford: Pergamon, pp. 42-67.

Kim, J & Mueller, C. W. (1978a). *Introduction to Factor Analysis: What It Is and How to Do It.*. London: Sage Publications.

Kim, J & Mueller, C. W. (1978b). *Factor Analysis Statistical Methods and Practical Issues*. London: Sage Publications.

Kolen, M. J. & Whitney, D. R. (1981). Comparison of four procedures for equating the tests of general educational development. Paper presented at the annual meeting of thee American Educational Research Association. Los Angeles, California.

Lokan, J., Ford, P. & Greenwood, L. (1996). *Maths & Science On the Line: Australian Junior Secondary Students' Performance in the Third International Mathematics and Science Study.* Melbourne: ACER.

Long, J. S. (1983). *Confirmatory Factor Analysis: A preface to LISREL*. Beverly Hills: Sage Publications.

Moss, J. D. (1982). *Towards Equality: Progress by Girls in Mathematics in Australian Secondary Schools.* Hawthorn, Victoria: ACER.

Norusis, M. J. & SPSS Inc (1993). *SPSS for Windows: Base System User's Guide: Release 6.0*. Chicago: SPSS Inc.

Peaker, G. F. (1969). How should national part scores be weighted? *International Review of Education,* 15, 229-237.

Rentz, R. R. & Bashaw, W. L. (1975*). Equating Reading tests with the Rasch model, Vol. I Final Report.* Athens, Georgia: University of Georgia: Educational Research Laboratory, College of Education.

Rosier, M. J. (1980). *Changes in Secondary School Mathematics in Australia.* Hawthorn, Victoria: ACER.

Smith, R. M. and Kramer, G. A. (1992). A comparison of two methods of test equating in the Rasch model. *Educational and Psychological Measurement,* 52 (4), 835-846.

Sontag, L. M. (1984). Vertical equating methods: A comparative study of their efficacy. DAI, 45-03B, page 1000.

Spearritt, D. (1994). Factor Analysis. In T. Husén & T.N. Postlethwaite (eds), *The International Encyclopedia of Education, (second ed.), vol.4.* Oxford: Pergamon, pp. 2230-2241.

Wright, B. D. (1995). 3PL or Rasch? *Rasch Measurement Transactions,* 9 (1), 408-409.

Wright, B. D., and Stone, M. H. (1979). Best Test Design: Rasch Measurement. Chicago: Mesa Press.

Willett, J. B. (1997). Change, Measurement of. In J.P. Keeves (ed.), *Educational Research, Methodology, and Measurement: An International Handbook, (second ed.)*, Oxford: Pergamon, pp. 327-334.

APPENDIX A

ERRORS OF ESTIMATIONS AND SCALING

In the present study, sources of errors of estimation and scaling that are related to the calculation of gains and losses in mathematics achievement are associated with the sampling design, the fitting of individual items to a scale based on the Rasch model (calibration) and the use of the equating constant based on the FIMS and TIMS data sets.

(1) The error associated with the sampling design for each data set was generated using WesVarPC computer program (Brick, *et al.*, 1997).

(2) The error associated with the use of the mean value of the equating constant arises from the equating using the nine common items in the FIMS and TIMS mathematics tests in the five state samples.

The error associated with the equating constant was estimated to be 0.104. The items employed for anchoring in the common item difference equating procedure are not a random sample of items but a fixed sample of specifically chosen items. Under these circumstances the error of the grand mean is given by $\frac{s}{\sqrt{n}}$ where n is the number of items in the sample for each state.

Standard error of equating constant = $\frac{0.699}{\sqrt{45}} = 0.104$

(3) For individual students the QUEST computer program (Adams and Khoo, 1993) provided a value for the magnitude of the measurement error associated with the estimation of student performance. This estimate for TIMS was about 35 scale units. In order to calculate the error arising from calibration the following formula was used:

$$\text{Standard error of calibration} = \text{Individual error } X \frac{1}{\sqrt{ESS}}$$

where ESS is effective sample size.

The effective sample size was approximately 229 for Australia TIMS sample and the calibration error for Australia was

$$35X\frac{1}{\sqrt{229}} = \frac{35}{15} = 2.3$$

Thus approximately two scale units are associated with this source of error

(4) The following formula was employed to calculate the total error .

$$se_{total} = \sqrt{(se_{64})^2 + (se_{94})^2 + (se_{Cal})^2 + (se_{equ})^2} \quad \text{where}$$

se_{64} = standard error associated with the estimated mean score of (1964 FIMS).
se_{94} = standard error associated with the estimated mean score of (1994 TIMS).
se_{Cal} = standard error associated with the use of calibration, and
se_{equ} = standard error associated with the use of the mean value of the equating constant.

APPENDIX B

EXTRAPOLATIONS OF ZERO AND PERFECT SCORES

The QUEST computer program (Adams and Khoo, 1993) by default does not process cases with perfect and zero scores, because both groups do not provide information for the calibration of the scale. Hence, in order to include those cases with perfect and zero scores in the calculation of the mean and standard deviation of the mathematics achievement test scores for each student sampled, the values of the perfect and zero scores were calculated by extrapolation from the logit table produced by the QUEST computer program. Table A shows the procedures employed to estimate the scores of cases with a perfect or zero score. The calculation of the scores of the FIMS students who had perfect and zero scores has been used here as an example. Table A1 shows the procedures employed to estimate the scores of cases with a perfect score. The first column indicates the top three raw scores (69, 68 and 67) excluding the highest possible raw score (70). The second column indicates the *logit* values obtained from the logit table generated by the QUEST computer program (Adams and Khoo, 1993). This column provides the Rasch scores corresponding to the top three possible raw scores in the test excluding the maximum score. D_1 gives the successive differences between the top three logit values. It was assumed that compared to the highest logit value, the perfect score was likely to be greater than a value equal to the difference between the top two scores and the difference between consecutive differences of the top three scores. Therefore, the following calculation was employed to estimate the perfect score. In order to get the first entry (0.73) in column D_1 the second highest logit value (4.17) was subtracted from the first highest logit value (4.90). The same procedure was applied to obtain the second entry (0.44) in which, the third highest value (3.73) was subtracted from the second highest value (4.17). The difference between the two entries in column D_1, that is the difference between 0.73 and 0.44, namely 0.29, was entered in column D_2. Therefore, the estimated Rasch score for the maximum raw score 70 is assigned in the column *Perfect Score* in Table A1. The score (5.92) was estimated by adding the highest logit value (4.90) for a score of 69 and the first entry in column D_1 (0.73) and the entry in column D_2 (0.29).

Table A. Estimation of Perfect and Zero Scores

Table A1. Estimation of Perfect Score

Scores	Logits	D_1	D_2	Perfect Score values
(max =70)				5.92
69	4.90			
68	4.17	0.73		
67	3.73	0.44	0.29	

Table A2. Estimation of Zero Score

Score	Logits	D_1	D_2	Zero Score
(max=70)				
3	-3.53			
2	-3.97	-0.44		
1	-4.70	-0.73	-0.29	
0				-5.72

D_1 : difference between adjacent logit values; D_2 : difference between adjacent values of D_1

For the estimation of the zero score it was assumed that compared to the lowest logit value, the zero score would most likely be less than the logit value for a score of one, by a value equal to the difference between the bottom two scores and the difference between consecutive differences of the bottom three scores. Hence, the same procedure was employed for the estimation of the zero score. However, the subtractions for the estimation of zero scores were from the bottom. Table A2 shows the estimation of the zero scores. Thus, to obtain the first entry (-0.73) in column D_1 the second lowest value (-3.97) was subtracted from the first lowest value (-4.70). In order to obtain the second entry (-0.44) in column D_1 the third lowest value (-3.53) was subtracted from the second lowest value (-3.97). Moreover, in order to obtain the entry in column D_2 the second lowest value of D_1 was subtracted from the lowest value of D_2. Therefore, the estimated Rasch score for the minimum raw score of zero is assigned in the column *Zero Score* in Table 2b. The score (-5.72) was estimated by adding the lowest logit value (-4.70) for a score of one and the lowest entry in column D_1 (-0.73) and the entry in column D_2 (-0.29).

International Education Journal Vol 1, No 1, 1999
http://iej.cjb.net

Native language interference in learning a second language: Exploratory case studies of native language interference with target language usage

Baljit Bhela
Flinders University School of Education

Introduction

The second language learning environment encompasses everything the language learner hears and sees in the new language. It may include a wide variety of situations such as exchanges in restaurants and stores, conversations with friends, reading street signs and newspapers, as well as classroom activities, or it may be very sparse, including only language classroom activities and a few books.

Regardless of the learning environment, the learner's goal is mastery of the target language. The learner begins the task of learning a second language from point zero (or close to it) and, through the steady accumulation of the mastered entities of the target language, eventually amasses them in quantities sufficient to constitute a particular level of proficiency (Dulay, Burt & Krashen, 1982 and Ellis, 1984).

This characterisation of language learning entails the successful mastery of steadily accumulating structural entities and organising this knowledge into coherent structures which lead to effective communication in the target language (Rutherford, 1987). If this is the case, then we would expect that well-formed accurate and complete target language structures would, one after another, emerge on the learner's path towards eventual mastery of the language. If the learner went on to master the language, we could, in principle, tabulate the expansion of his/her repertoire up to the point where all of the well-formed structures of the target language had been accounted for (Beardsmore, 1982 and Hoffman, 1991).

In reality this is not the case. Second language learners appear to accumulate structural entities of the target language but demonstrate difficulty in organising this knowledge into appropriate, coherent structures. There appears to be a significant gap between the accumulation and the organisation of the knowledge. This then raises a critical question - what kinds of language do second language learners produce in speaking and writing? When writing or speaking the target language (L2), second language learners tend to rely on their native language (L1) structures to produce a response. If the structures of the two languages are distinctly different, then one could expect a relatively high frequency of errors to occur in L2, thus indicating an interference of L1 on L2 (Dechert, 1983 and Ellis, 1997).

Previous Research and the Importance of this Research

Extensive research has already been done in the area of native language interference on the target language. Dulay et al (1982) define interference as the automatic transfer, due to habit, of the surface structure of the first language onto the surface of the target language. Lott (1983: 256) defines interference as 'errors in the learner's use of the foreign language that can be traced back to the mother tongue'.

Ellis (1997: 51) refers to interference as ‘transfer’, which he says is 'the influence that the learner’s L1 exerts over the acquisition of an L2'. He argues that transfer is governed by learners’ perceptions about what is transferable and by their stage of development in L2 learning. In learning a target language, learners construct their own interim rules (Selinker, 1971, Seligar, 1988 and Ellis, 1997) with the use of their L1 knowledge, but only when they believe it will help them in the learning task or when they have become sufficiently proficient in the L2 for transfer to be possible.

Ellis (1997) raises the need to distinguish between errors and mistakes and makes an important distinction between the two. He says that errors reflect gaps in the learner’s knowledge; they occur because the learner does not know what is correct. Mistakes reflect occasional lapses in performance; they occur because, in a particular instance, the learner is unable to perform what he or she knows.

It appears to be much more difficult for an adult to learn a second language system that is as well learned as the first language. Typically, a person learns a second language partly in terms of the kinds of meanings already learned in the first language (Carroll, 1964; Albert & Obler, 1978 and Larson-Freeman & Long, 1991). Beebe (1988) suggests that in learning a second language, L1 responses are grafted on to L2 responses, and both are made to a common set of meaning responses. Other things being equal, the learner is less fluent in L2, and the kinds of expressions he/she uses in L2 bear telltale traces of the structure of L1.

Carroll (1964) argues that the circumstances of learning a second language are like those of a mother tongue. Sometimes there are interferences and occasionally responses from one language system will intrude into speech in the other language. It appears that learning is most successful when the situations in which the two languages (L1 and L2) are learned, are kept as distinct as possible (Faerch and Kasper, 1983). To successfully learn L2 requires the L2 learner to often preclude the L1 structures from the L2 learning process, if the structures of the two languages are distinctly different.

Beardsmore (1982) suggests that many of the difficulties a second language learner has with the phonology, vocabulary and grammar of L2 are due to the interference of habits from L1. The formal elements of L1 are used within the context of L2, resulting in errors in L2, as the structures of the languages, L1 and L2 are different.

The relationship between the two languages must then be considered. Albert and Obler (1978) claim that people show more lexical interference on similar items. So it may follow that languages with more similar structures (eg English and French) are more susceptible to mutual interference than languages with fewer similar features (eg English and Japanese). On the other hand, we might also expect more learning difficulties, and thus more likelihood of performance interference at those points in L2 which are more distant from L1, as the learner would find it difficult to learn and understand a completely new and different usage. Hence the learner would resort to L1 structures for help (Selinker, 1979; Dulay et al, 1982; Blum-Kulka & Levenston, 1983; Faerch & Kasper, 1983, Bialystok, 1990 and Dordick, 1996).

Dechert (1983) suggests that the further apart the two languages are structurally, the higher the instances of errors made in L2 which bear traces of L1 structures. In both cases the interference may result from a strategy on the part of the learner which assumes or predicts equivalence, both formally and functionally, of two items or rules sharing either function or form. More advanced learning of L2 may involve a greater number of rules or marking features for distinguishing between the two languages. This then raises a pertinent question - does the L2 text have to be syntactically correct for its meaning to be understood? Do the identified errors in the written text reduce semantic and syntactic acceptability? The answer lies in several domains: the L2 learner's purpose in learning the target language, the learner's L2 proficiency level of the target language and the knowledge state of the learner in L1 and L2.

The focus of the case studies is on specific instances of L1 interference on L2 in the syntactic structures of the second language learner's writing. The present study also identifies the effect of the differences and/or similarities between the structures of L1 and L2 on the target language.

The case studies concentrate on the effect of each of the areas of difficulty identified on a native speaker's interpretation of the written text. The case studies also identify the importance of the learner's knowledge of the syntactic structures of L1, which cause difficulty in L2. With this knowledge the learner is made aware of the errors made and how they may be rectified. This aspect of the study also provides new information in the L2 learning context. Last but not least, the case studies identify the language use and the knowledge of the learner. Hence this study attempts to provide up-to-date evidence in the current L2 learning context.

An important aspect of this study is that it provides an interesting comparison of four languages, namely Vietnamese, Cambodian, Spanish and Italian. The combination of two Asian and two European languages is a move away from a previous research focus on mainly the European languages and this is useful for the current local teaching context.

Research Questions

The case studies were designed to answer the following questions:

1. Are there differences and/or similarities between the syntactic structures of L1 and L2 in a written task in each of the cases?

1a. What are the instances where the syntactic structure of L1 is used in L2, causing an error?

1b. What are the instances where the absence of a syntactic structure in L1 creates a difficulty for the learner in L2?

2. What is the effect of each of the noted areas of difficulty on interpretation of meaning by a native speaker of English?

3. What is the learner's knowledge of the syntactic structure of L1, which causes difficulty in L2?

4. What is the learner's knowledge of the syntactic structure of L2?

The research scope of this paper is limited to the analysis of writing samples of four adult second language learners in the language classroom, with a focus on syntactic structures and takes into account errors made in semantics and spelling.

Research Methodology

The case study methodology in this study was not an experimental intervention. It was designed to uncover something of the complexity of language use in a particular sample of language learners and so it had an explicit descriptive purpose. It aimed to analyse the use of specific parts of language and to use the results of that analysis to make judgements about the status of the L1-L2 interference hypothesis. The interview was a flexible procedure that allowed for probing of the participants' linguistic knowledge. The research questions posed were mainly "what" questions that were exploratory. A goal of the study was to develop a pertinent hypothesis and propositions for further inquiry.

Participants

There were four participants in the study - a Spanish-speaking 21-year-old female (Bianca), a Vietnamese-speaking 39-year-old female (Cath), a Cambodian-speaking 50-year-old female (Sabi) and an Italian-speaking 65-year-old male (Mato). Writing is important for these learners as they either have young school-aged children or grandchildren who request some help with schoolwork from time to time.

Tasks

The four learners were given two sets of sequential pictures, one at a time, and asked to write a story beginning with the first picture and ending with the last, in the order presented in each set. The first set of pictures related to a boy deciding to play tennis instead of washing the car and the second set of pictures related to a driver driving in the opposite direction of the traffic. There was no time limit for the task. However they had to ensure that they had a logical sequence in the written story which related to the pictures.

The learners were then asked to write the same story a second time, in the native language. They were then asked to write a second story in English and the native language for the second set of sequential pictures. The learners were asked to attempt the tasks individually without any group interaction initially. After an individual attempt, they were allowed to interact with each other if they wished. The tasks were part of the classroom activities done in the presence of the teacher.

Writing two stories each in English as well as the native language provided a broader base for the analysis of the errors made. It also provided a suitable sample of written performance, thus allowing a more reliable estimate of the participants' competence.

Interview

After the writing tasks, the four learners were interviewed individually, which were tape-recorded where they were asked to explain why and how they used a specific L1 or L2 structure if there was an error identified. They were also asked what they knew about the structures of L1 and L2 and to make judgements of semantic acceptability of sentences in L1 and L2. They were then asked to self-correct identified errors in the L2 text.

Analysis Procedures

The analysis of the learners’ L1 written texts was done with the help of native language experts, while I analysed the English texts. Three L2 native speaker teachers were asked to interpret the learners’ L2 written texts and rate these texts for semantic and syntactic acceptability. The purpose here was to answer the pertinent question raised - does the L2 text have to be syntactically correct for its meaning to be understood for L2 learners at the assessed level of L2 proficiency?

Results

The four subjects completed the tasks in an hour. The non-segmented and unedited versions of each learner's written texts were analysed.

L1 and L2 proficiency levels

The four learners were assessed before the tasks, using the Australian Second Language Proficiency Ratings (ASLPR) to determine their L2 writing proficiency. The ASLPR has 12 proficiency levels within a scale of 0 to 5, with 5 indicating native-like proficiency. The ASLPR was also used as a comparison gauge by the native language teachers to identify the learners' L1 writing proficiency in comparison to their L2 proficiency. These four learners were at level 1+ in their writing skills according to the ASLPR. According to Wylie and Ingram (1995) learners at this level can write simple social correspondence; their language is creative enough to use stock phrases and complex enough to convey in a simple way, their own attitudes to familiar things; they make several mistakes but generally get their ideas across. Table 1 shows the learners' L1 and L2 proficiency levels. The learners were found to have similar levels (1+) in their L1 and L2.

Table 1. Proficiency levels of L1 and L2

Learners' proficiency	Sabi	Cath	Mato	Bianca
L1	1+	1+	1+	1+
L2	1+	1+	1+	1+

L1 errors

An analysis of each learner's writing indicated several grammatical errors. Table 2 shows the L1 errors made by the learners in their written texts, a 'x' denotes an error/s and a '_' denotes a correct response/s made with the specific structure, while a '0' denotes an absent L1 structure altogether.

Table 2. L1 errors for all learners

L1 errors	Sabi	Cath	Mato	Bianca
Apostrophe	0	0	0	x
Punctuation	x	x	x	x
Spelling	x	x	x	x
Prepositions	x	x	x	x
Capital letters	x	x	x	x
Present & past continuous tenses	√	√	√	x
Subject pronouns	√	√	√	x
Vocabulary	x	√	√	x
Passive & active voice	0	0	√	x

L2 errors

Table 3 shows the L2 errors made by the learners in the two writing tasks. For example, all four learners made errors in the use of punctuation (denoted by a 'x'). Mato and Bianca did not use the repeated pronoun as this structure is absent in their L1 (denoted by a '0'). Bianca used subject pronouns appropriately in her L2 texts (denoted by a '_').

A comparison of the analyses of L1 and L2 showed eight syntactical areas bearing signs of direct interference of L1 on L2. The results are shown in Table 4 below, which shows the errors made by the four learners in both their L1 and L2 texts, where the L1 errors were transferred to the L2 texts.

Pairs of languages

Table 5 shows the differences and similarities between the syntactical structures of the learners' native languages when compared to English. The 'A' denotes an absent structure, the 'P' denotes an existing structure with limited use in L1 and the 'S' denotes a similar structure to English. The four languages may be divided into pairs as some of the structures of Vietnamese/ Cambodian and Italian/Spanish bear similarities.

Table 3. L2 errors for all learners

L2 errors	Sabi	Cath	Mato	Bianca
Apostrophe	x	x	x	x
Contractions	√	√	√	x
Punctuation	x	x	x	x
Articles	x	x	√	√
Prepositions	x	x	x	x
Spelling	x	x	x	x
Capital letters	x	x	x	x
Repeated pronouns	x	x	0	0
Subject pronouns	x	x	x	√
Present & past continuous tenses	x	x	x	x
Past tense	x	x	x	√
Adverbs	x	x	√	x
Plurals	x	√	√	√
Incomplete sentences	x	x	√	√
Vocabulary	x	x	x	x
Passive & active voice	x	x	x	x

Table 4.
A comparison of the L1 and L2 analyses indicating areas of L1 interference on L2

L2 errors	Sabi	Cath	Mato	Bianca
Possessive apostrophe	x	x	x	x
Punctuation	x	x	x	x
Passive & active voice	x	x	√	x
Prepositions	x	x	x	x
Spelling	x	x	x	x
Capital letters	x	x	x	x
Repeated pronouns	x	x	x	x
Present & past continuous tenses	x	x	x	x

Table 5. Pairs of languages

	Pairs of languages							
L1 structures	**Cambo**		**Viet**		**Italian**		**Spanish**	
Possessive apostrophe	A	x	A	x	A	x	A	x
Punctuation	S	x	S	x	S	x	S	x
Passive voice	P	x	P	x	S	√	S	x
Prepositions	P	x	S	x	P	x	P	x
Repeated pronouns	P	x	P	x	A	x	A	x
Present & past continuous tenses	P	x	P	x	S	x	S	x
Capital letters	S	x	S	x	P	x	P	x

The results from Table 4 indicating the learners' L2 errors are also shown alongside each structure. For example, the 'x' denotes an error/s made in L2 and a '_' denotes a correct use of the structure. Table 5 then shows that although a structure is present in L1, the learners still made errors with its use in L2, indicating a lack of understanding of its L2 use and the learners used the L1 form in L2, making errors in L2. Where a structure is absent in L1, for example the possessive apostrophe, the learners did not understand its use in L2, once again making errors. Where the structure, for example the use of punctuation, is similar in its use in L1 and L2, the learners made errors with its use, as they have also made similar errors with its use in L1 in all the cases. This indicates direct interference of L1 on L2.

Self-editing

The learners were observed in terms of the oral group interaction during the tasks. They asked each other for help, particularly with spelling and vocabulary. In the individual interviews, they were asked to explain why they had used specific L1 and L2 structures.

Table 6 shows the L2 self-editing instances by each of the learners. All four learners were able to self-edit some of the errors made in their L2 texts after these errors were pointed out to them individually. In the self-editing, the learners concentrated on the correction of spelling and there were a few instances of syntactical error correction such as punctuation.

Table 6. L2 self-editing

Learners	Sabi	Cath	Mato	Bianca
L2 self-editing	√	√	√	√

L2 semantic acceptability

Three L2 native speaker teachers (NST) were asked to interpret the learners' L2 written texts without showing them the sequential sets of pictures the learners had been given for the tasks. The teachers were asked to rate the texts on a scale of 1 (poor), 2 (average) and 3 (good) for semantic and syntactic acceptability in terms of the stories written for the sequential sets of pictures, as indicated in Table 7 below. Text 1 and Text 2 were given similar ratings by the three native speaker teachers. This then meant that the type of writing each learner produced in both texts, Text 1 and Text 2 were of a similar level and could be understood by the L2 native speaker teachers, despite the errors found in the texts. The pertinent question previously raised in this paper was answered - the L2 text does not have to be syntactically correct (by L2 standards) for its meaning to be understood, for L2 learners at the assessed level of L2 proficiency.

Table 7. Native speaker teachers' rating for L2 semantic acceptability

	Rating for L2 semantic acceptability					
	Text 1			**Text 2**		
Learners	**NST 1**	**NST 2**	**NST 3**	**NST 1**	**NST 2**	**NST 3**
Sabi	2	2	2	2	2	2
Cath	3	3	3	3	3	3
Mato	3	3	2	3	3	2
Bianca	2	2	2	2	2	2

General discussion

This study provided a view and an indication of the kinds of language second language learners produced in writing tasks in the classroom. It also supplied evidence of L1 interference with L2, its extent and effects, as shown in the analysis of the learners' written L1 and L2 texts. This was clearly shown in the way that the learners used their L1 structures to help them form their L2 texts, indicating a direct interference of L1 on L2.

The four learners have received native language linguistic input from their individual environments and positive reinforcements for their correct repetitions and imitations. As a result, habits have been formed which have influenced the L2 learning process as these learners have started learning L2 with the habits associated with L1. These habits interfere with those needed for L2 learning, and new habits are formed. The errors made in L2 are thus seen as L1 habits interfering with the acquisition of L2 habits (Beebe, 1988 and Seliger, 1988). This theory also propounds the idea that where there are similarities between L1 and L2, the learners use L2 structures with ease; where there are differences, the learners have difficulty. The four learners have constructed their own L2 interim rules with the use of their L1 knowledge to help them in the writing tasks, resulting in L2 errors (Ellis, 1997).

Some L2 errors identified in Table 3 such as articles, adverbs, past tense, plurals, contractions and incomplete sentences, were not included in the discussion of the L1 - L2 interference. This was because these errors did not appear in the L1 texts, thus indicating that, although the learners made these errors in their L2 texts, the structures were used appropriately in the L1 texts or the learners did not use these structures at all as these were absent structures in the L1.

The four learners appear to find it difficult to use appropriate L2 responses that are as well formed as their L1 structures. They use the L2 structures partly in terms of the structures already learned in their L1. Hence their L1 responses are grafted on the L2 responses and the kinds of L2 expressions used bear tell-tale traces of the L1 structures (Larson-Freeman & Long, 1991 and Ellis, 1997). This was clearly shown in the way that the learners used L2 structures such as punctuation, capital letters, prepositions and the present and past continuous tenses in their L2 texts. They found these structures difficult to use as these structures are used in a different form in their L1. In some instances, an absent L1 structure such as the apostrophe and the active and passive voice, caused a difficulty for the learners as they were unfamiliar with its use in L2, resulting in errors which reflect a gap in the learners' knowledge (Ellis, 1997).

As Dechert (1988) has already suggested, the further apart L1 and L2 are structurally, the higher the instances of errors made in L2 which bear traces of L1 structures. An important outcome of this study is the significance of the effect of the differences between the structures of L1 and L2 on the L2 written text. Given the proficiency level of the learners in the study, the learners' L2 texts remain semantically acceptable by L2 teachers as shown in the analysis. This then means that the L2 texts do not have to be syntactically correct for its meaning to be understood. The identified L2 errors do not reduce the semantic acceptability of the L2 texts.

Does the learner have to “think” in the target language to be able to produce a meaningful response which may not be syntactically correct but which may still be understood and semantically acceptable? The answer to this question poses a major implication in the second language classroom. If the learner is able to write a semantically acceptable text in L2 (according to L2 standards), then correct syntax need not be the focus of classroom instruction, given the existing knowledge base of the learner whose main purpose of learning L2 is to communicate information in a meaningful way.

This has implications for the teaching and learning process. An understanding of the L1 syntactical structure and the type of errors made in L2 as well as the extent of the learner’s knowledge of L1 and L2 syntactical structures, will assist the teaching and learning process by allowing an individualised learning program for each learner. The teacher will be able to predict possible future errors in the target language and may begin to attribute a cause to an error with some degrees of precision. The teacher can also build up a picture of the frequency of types of errors; thus it would be possible to find out whether, for example, L1 interference, or teaching techniques, or problems inherent in L2, are the major cause of the learner's errors. In this way it is possible to plan classes giving very specific help to the learners.

This case study then paves the way for future research in other areas of second language teaching and learning. Last but not least, this study contributes significantly to the base of knowledge in the second language learning and teaching literature on the effects of interference of L1 on L2.

Conclusions

The major concern of this paper has been with the observable features of interference of L1 on L2 and what its effects are on the syntactic structure of a written task of a second language learner. The learners have used some L1 structures to produce appropriate responses in L2, producing semantically acceptable texts. Subsequently, the learners have also used L1 structures interchangeably with L2 structures, producing inappropriate L2 responses, indicating an interference of L1 on L2. These structures are used to make them understood and reflect the way they arrive at a certain usage at a specific point (Faerch & Kasper, 1983). These structures do not reflect failure in any way but are a means to increase their resources in order to realise their communicative intentions. In using the L1 structures, the learners have taken some risks that include guessing of a more or less informed kind. They have attempted to use invented or borrowed items, all more or less approximated to the rules of L2 structure as far as their knowledge of L2 allows.

When the learners experience gaps in their L2 syntactical structures, they adjust the form of their L2 written responses by using syntactical items which are part of their L1. The analysis of the learners' writing revealed the extent to which their L2 responses are affected by their L1, the procedures used to express concepts for which L2 syntax is unknown and the extent to which and the manner in which L1 syntax interferes with L2 (Bialystok, 1990). The L2 errors made are traceable to the learners' L1 and we can conclude that there is definite interference of L1 on L2 as indicated in the analysis of the eight syntactical areas discussed.

The four learners relate L2 syntax to what they already know about language. The most salient facts they possess about language are those of L1. In the process of attempting to relate L2 to L1, they speculate about the similarity or difference between L2 and L1. The result is a subsumption of L2 under known categories in L1 competence and hence a translation process has taken place (Seligar, 1988). Where the structures of L1 and L2 are similar, the learner' lack of understanding its use in L1 is also reflected as an error in L2.

The use of L1 structures as a principle of fundamental language organisation and processing has immediate serviceability for these learners. The learners bring the form and meaning of both L1 and L2 into closer alignment and thus render usable a complex portion of L2 syntax that would otherwise be for the time being, inaccessible to them. The prior disposition of L1 has affected the L2 responses.

Blum-Kulka and Levenston (1983) contend that all second language learners begin by assuming that for every word in L1 there is a single translation equivalent in L2. The assumption of word-for-word translation equivalence or 'thinking in the mother tongue (L1)' is the only way a learner can begin to communicate in a second language.

This has been clearly indicated in this study where the second language learners have adopted their L1 structures to help them in their L2 texts. These learners will not attain mastery of the target language as long as the process of translation equivalence is in place. Blum-Kulka and Levenston assert that mastery of the second language involves the gradual abandonment of the translation equivalence, the internalisation of the syntactical structures in L2 independently of the L1 equivalent, and the ability to 'think in the second language'.

These learners have accumulated structural entities of L2 but demonstrate difficulty in organising this knowledge into appropriate, coherent structures. There is a significant gap between the accumulation and organisation of this knowledge. When writing in the target language, these learners rely on their native language structures to produce a response, as shown in this study. As the structures of L1 and L2 have differences, there has been a relatively high frequency of errors occurring in the target language, thus indicating an interference of the native language on the target language, as expected.

Limitations of the Study

This case study was based on an observation of four adult second language learners and an analysis of each of their writing tasks in the classroom. As such, the sample involved was small and there was a limited range of languages analysed - Spanish, Italian, Vietnamese and Cambodian. This being the case, no generalisations for all second language learners are made. The value of this study is, paradoxically, its generalisability to a similar set of circumstances for the type of learners identified in the study. It is generalisable to theoretical propositions and not to populations.

References

Albert, M.L. & Obler, L.K. 1978, *The Bilingual Brain: Neuropsychological and Neurolinguistic Aspects of Bilingualism*, Academic press, New York.

Beardsmore, H.B. 1982, *Bilingualism: Basic Principles*, Tieto, Avon.

Beebe, L.M. (ed.) 1988, *Issues in Second Language Acquisition: Multiple Perspectives*, Newbury, London.

Bialystok, E. 1990, *Communication Strategies: A Psychological Analysis of Second Language Use*, Basil Blackwell, Oxford.

Blum-Kulka, S. & Levenston, E.A. 1983, 'Universals of lexical simplification' in *Strategies in Interlanguage Communication*, eds. C. Faerch and G. Kasper, Longman, London.

Carroll, J.B. 1964, *Language and Thought*, Prentice-Hall, Englewood Cliffs.

Dechert, H.W. 1983, 'How a story is done in a second language' in *Strategies in Interlanguage Communication*, eds. C. Faerch and G. Kasper, Longman, London.

Dordick, M. 1996, Testing for a hierarchy of the communicative interference value of ESL errors, *System*, vol. 24, pp 299-308.

Dulay, H., Burt, M. & Krashen, S. 1982, *Language Two*, Oxford University Press, New York.

Ellis, R. 1984, *Classroom Second Language Development: a study of classroom interaction and language acquisition*, Pergamon Press, Oxford.

Ellis, R. 1997, *Second Language Acquisition*, Oxford University Press, Oxford.

Faerch, C. & Kasper, G. 1983, 'Plans and strategies in foreign language communication', in *Strategies in Interlanguage Communication*, ed. C. Faerch and G. Kasper, Longman, London.

Hoffman, C. 1991, *An Introduction to Bilingualism*, Longman, London.

Larson-Freeman, D. & Long, M.H. 1991, *An Introduction to Second Language Acquisition Research*, Longman, New York.

Lott, D.1983, Analysing and counteracting interference errors, *ELT Journal*, vol.37/3, pp 256- 261.

Rutherford, W.E. 1987, *Second Language Grammar: Learning and Teaching*, Longman, London.

Seligar, H. 1988, 'Psycholinguistic Issues in Second Language Acquisition' *in Issues in Second Language Acquisition: Multiple Perspectives*, ed. L.M. Beebe, Newbury, London.

Selinker, L. 1971, 'The psychologically relevant data of second language learning' in *The Psychology of Second Language Learning*, ed. P. Pimsleur and T. Quinn, Cambridge University Press, London.

Wylie, E. & Ingram, D.E. 1995, *Australian Second Language Proficiency Ratings: General proficiency version for English*, Griffith University, Queensland.

A Cross-Cultural Comparison of Student Concerns in the Teaching Practicum

Rosalind Murray-Harvey
roz.murray-harvey@flinders.edu.au

Halia Silins
halia.silins@flinders.edu.au

Judith Saebel
judith.saebel@flinders.edu.au

School of Education, Flinders University of South Australia

There is general consensus in the literature that students consider the practicum to be a highly valued component of their teacher education degree. Nevertheless, there are wide ranging concerns reported by students related to their teaching practice. This paper reports on these concerns in the form of a cross-cultural comparison of an Australian and a Singaporean sample of students.

Singaporean and Australian students completing their first practicum independently responded to a questionnaire based on the Survey of Practicum Stresses (D'Rozario & Wong, 1996). The psychometric properties of their 7-factor model were tested using the Australian data. This resulted in a 4-factor model, which was confirmed using structural equation procedures. Details of effective but under-employed analysis techniques are presented. This model was employed subsequently to provide cross-cultural comparisons of student concerns in the teaching practicum. Significant differences between the stresses experienced by Singaporean and Australian students point to the need to understand student stress within a cultural context.

Introduction

Stress experienced by students in their practicum has been reported in enough studies to indicate that it is not an isolated phenomenon. In order to maximize the benefits of the teaching practicum for student teachers and for teacher educators, both need to address the concerns of students related to their teaching practice experiences.

MacDonald (1993), along with other researchers into student teacher stress (Campbell-Evans & Maloney, 1995; Capel, 1997; D'Rozario & Wong, 1996; Elkerton, 1984; Morton, Vesco, Williams, & Awender, 1997), confirm that while students regard the teaching practicum as a valuable, if not the most valued, part of their teacher education program, they also consider it to be the most stressful. The significance of identifying sources of student teacher stress lies in the evidence that stress affects teacher behaviour and this in turn reduces classroom effectiveness, particularly in relation to effects of lower pupil achievement and increased levels of pupil anxiety. Elkerton (1984) exhorted teacher educators to identify stresses associated with the practicum and to assist students to effectively manage these stresses. Morton et al. (1997) pointed to the need to change the nature of the role of teacher and university supervisors from a more directive to a more collaborative one in order to reduce student stress related to evaluation and assessment. Jeans and Forth (1995) also drew attention to the need to bridge the worlds of theory and practice in the design and implementation of pre-service teacher education programs.

MacDonald's (1993) research identified that sources of stress were mainly generated by inconsistencies in the way students were evaluated by teachers, varying expectations of student performance and conformity between teachers, and marked variations in the quality of feedback given to students by their supervising teachers. Gender emerged as an issue in research conducted by D'Rozario and Wong (1996) with student teachers in Singapore, and by Morton et al. (1997). It was reported in both studies that females generally find the practicum experience more stressful than males. At a more general level, Bowers, Eichner, and Sacks (1982) suggested that teacher preparation had not paid enough attention to the psychological 'readiness' of student teachers by concentrating more on methodology and less on preparing students to cope with the inevitable anxieties and stresses associated with students' roles, relationships and responsibilities of teaching.

In the literature on the practicum that reports on student teacher concerns, stresses, and anxieties, only Morton et al. (1997) were found to have taken a cross-cultural focus. They noted that differential reactions to stressors are likely to be a function of variables such as personality, sex, and culture. "Thus male and female student teachers may respond differently to the specific stressors of the teaching experience. Similarly, student teachers in one country may differ in perceived stressors from student teachers in another country" (p. 70). These authors posited that variables including teacher-status, teacher income, teacher demand, and teacher stress could account for differences between cultures in student teachers' cognitive appraisals of, and anxieties about, their school experience. Their cross-cultural research involved a factor analysis of data from Canadian student teachers who completed Hart's Student Teacher Anxiety Scale (STAS) to compare the anxieties of the Canadian students with the British students who provided the data for Hart's original factor analysis. Hart's analysis produced four anxiety factors which were labelled *Evaluation Anxiety*, *Pupil and Professional Concerns*, *Class Control*, and *Teaching Practice Requirement*s. Morton et al. (1997) also reported a 4-factor solution; their factors were called *Evaluation Anxiety*, *Pedagogical Anxiety*, *Classroom Management Anxiety*, and *Staff Relations Anxiety*. Both Canadian and British students were most anxious about evaluation. Given the many features common to the two cultures (language, history, curriculum, politics), the observed similarities between the groups were hardly surprising. However, the writers were not prepared for the finding that "evaluation anxiety appears to be paramount regardless of country" (p.72).

During the analysis of data in another study on student teacher stress during the practicum (Murray-Harvey, Slee, Lawson, Silins, Banfield, & Russell, 1999) that employed the Survey of Practicum Stresses (D'Rozario & Wong, 1996), marked differences emerged between the practicum concerns of teacher education students in Singapore and Australia. This led us to research the concerns about the practicum held by students in other different cultural contexts. We asked: (1) How are Singaporean and Australian students' concerns about the practicum conceptualized? and (2) What concerns teacher education students most and least in their practicum? In this study stress was understood to involve the students' perceptions of demands on them (expressed concerns) associated with the teaching practicum.

The Cross-Cultural Context

The Singaporean Context

Compared with Australia, Singapore has a highly centralized system of education. "Schools are for the teaching of a national curriculum to pupils, not for reducing or solving society's ills, or achieving gender equality…" (Wong, Chiew, Gopinathan, & D'Rozario, 1998, p. 34). Pupils in Singapore's primary schools are streamed at grades 4 and 6 into different curriculum tracks according to ability. They also sit a national examination for promotion from primary to secondary school. Most schools are coeducational and most are neighbourhood schools. Class sizes are comparatively large (35-44 pupils). Overall, this means that student teachers work in schools that are relatively homogenous. The curriculum is prescribed and timetables are fixed.

The Australian Context

In Australia, despite efforts towards a national curriculum, individualism persists and even within the Australian States, schools vary in the ways they work with curriculum guidelines to reflect their own community's particular needs. Thus the context in which student teachers practice is likely to be much more variable than it is for Singaporean students. Adaptability and flexibility are regarded by the cooperating teachers as positive attributes of their student teachers. Teachers need to manage a continually changing timetable; pupil movement in and out of the classroom to attend, for example, specialized music, enrichment, specific needs, or language programs; and within the classroom – parent involvement (especially in the early years), and visitors. Class sizes are typically around 30 pupils but are increasing. As in Singapore, primary children generally attend a local, coeducational school.

Wong et al. (1998) suggested that teaching in Singapore is not a high status occupation so attracting capable entrants is difficult. Likewise, in Australia teaching does not command high status. Entrants to Education degrees generally achieve scores comparable with those of the generalist BA and BSc degrees. Similarly in Singapore and Australia, other specialized qualifications "enjoy higher rates of return" (p. 39). Another common feature of the two cultures is the relatively stable political environment. Probably the greatest contrast between the environments in which our student teachers practice their teaching is the more formal, centralized, and regulated system in Singapore.

Teacher Education in Singapore

Students at the National Institute of Education in the Diploma in Education (General) are graduates enrolled in a two-year program that prepares them to teach in primary schools. The BA/BSc with Diploma in Education is a four-year undergraduate program for teaching in both primary and secondary schools. Students in both programs undertake their first practicum for five weeks in a primary school. Assessment is non-graded (Pass/Fail).

Teacher Education in Australia

There are over 35 schools of teacher education in Australia and there is variation among them. Flinders University in Adelaide, South Australia, offers undergraduate and graduate-entry Bachelor of Education degrees. The undergraduate degrees prepare students to teach in primary schools (Reception to Year 7) or in middle schools (upper primary through junior secondary - years 6 to 10).These four-year degrees also admit graduate-entry students. There are also specialist Secondary teaching and Special Education graduate-entry programs. The professional development component of the teacher education program, which is concentrated into the final two years (four semesters) of the 4-year undergraduate degree represents the full BEd program for graduate students who enter with a completed university degree. Students in all programs receive a non-graded assessment (Satisfactory/Not Yet Satisfactory). The first practicum is a 6-week teaching block in Semester 2 in the same school that students visited for 2 weeks earlier in Semester 1.

The Sample

The Singaporean and Australian samples were similar in many respects. They included students from both the undergraduate 4-year programs and the graduate-entry 2-year programs. Both samples represented students who completed the Survey of Practicum Stresses after their first practicum experience. The Singaporean and Australian samples comprised 397 and 309 students, respectively. In the Singaporean sample there were more females than males (13.6%) and the majority of students were placed in government schools (78.6%). Males were typically under-represented in the Australian sample (26.3%) as well. The majority of students undertook their practicum in government schools (90.4%). All were placed in a primary school with the exception of ten secondary student teachers who were placed in secondary schools. There were no age data for the Singaporean sample. For the Australian sample, students' ages ranged from 20 to 53 years (M = 26 years).

The Survey Instrument

The Survey of Practicum Stresses (SPS) was renamed the Perceptions of Teaching questionnaire for the Australian students on request of the University ethics committee. D'Rozario and Wong (1996) developed the SPS to examine areas of stress experienced by first year teacher education students in Singapore. The same 29-item questionnaire was administered to the Australian students. The questionnaire consists of items representing experiences related to the practicum that students may find stressful, for example: managing the class and enforcing discipline; coping with the overall workload; being evaluated by the supervisor; and, fear of failing the practicum. Students' responses indicate how often the experience may have stressed them on a 4-point Likert scale, where 1 = *Never Stressed Me*, 2 = *Stressed Me Some of the Time*, 3 = *Stressed Me Most of the Time*, and 4 = *Stressed Me All the Time*. The possible range of responses to the 29 items on the questionnaire is from a minimum score of 29, indicating that the student experienced no stress on any item (29 x score of 1) to a maximum score of 116, indicating that the student was always stressed (29 x score of 4).

The comparisons presented between Singaporean and Australian students are based on the statistics reported by D'Rozario and Wong (1996). The raw data were not available.

How Are Singaporean and Australian Students' Concerns Conceptualized?

D'Rozario and Wong (1996) used the data from 397 first year student teachers in Singapore to explore the psychometric properties of the SPS. Employing exploratory factor analysis with principal component extraction and Varimax rotation, D'Rozario and Wong evolved a 7-factor model. This then lead to the creation of the following seven SPS subscales: *Overall Performance*, *Workload*, *New Colleagues*, *Cooperating Teacher*, *Supervisor*, *Teaching and Managing*, and *Helping.*

To understand how Australian students' concerns are conceptualized, we applied the same factor analytic procedure to the 309 responses collected from Australian teacher education students. The best fitting model that emerged from the analysis of the Australian data (for Practicum 1) was a 4-factor model, not a 7-factor model like in the D'Rozario and Wong's 1996 study. The new factors (and the resulting subscales) were labelled as *Teaching*, *Preparation*, *University Evaluation*, and *School Evaluation*. Their item composition is given in Table 1. Cronbach alphas for individual subscales indicated good reliabilities (Teaching = .85; Preparation = .77; University Evaluation = .85; School Evaluation = .74). This 4-factor model was subsequently tested applying confirmatory factor analysis via structural equation modelling (SEM), carried out using LISREL 8.12a (Jöreskog & Sörbom, 1993).

Table 1. Item Composition of the 4-factor Model of Australian Student Teachers' Concerns

New Subscale	SPS Item (SPS Subscale[a]/SPS Item Number)
Teaching	Managing the class and enforcing discipline (TM/24)
	Delivering the lesson (TM/19)
	Managing groupwork (TM/22)
	Managing the individual seatwork (TM/23)
	Establishing rapport with pupils (TM/18)
	Giving appropriate feedback to pupils (TM/21)
	Marking pupils' written work (WK/28)
	Teaching mixed ability classes (HL/27)
	Helping pupils with learning difficulties (HL/25)
	Helping pupils with emotional/behavioural problems (HL/26)
	Communicating concepts to pupils (TM/20)
	Having high expectations of my teaching performance (OP/3)
Preparation	Overall teaching workload (WK/5)
	Writing detailed lesson plans (WK/15)
	Managing time (WK/29)
	Striking balance between practicum and personal commitments (OP/2)
	Selecting appropriate content for my lessons (WK/16)
	Preparing resources for my lessons (WK/17)
	Others expecting me to perform tasks beyond my competency (OP/4)
	Managing practicum-related assignments (OP/6)
University Evaluation	Being observed by my supervisor (SU/13)
	Being evaluated by my supervisor (SU/14)
	Communicating with and relating to my supervisor (SU/12)
School Evaluation	Being observed by my Cooperating teacher(s) (CT/10)
	Being evaluated by my Cooperating teacher(s) (CT/11)
	Communicating with/relating to my Cooperating teacher(s) (NC/9)
	Fear of failing the practicum (OP/1)
	Communicating with/relating to teachers in the school (NC/8)
	Communicating with/relating to Principal/Vice-Principal (NC/7)

Note. [a] OP = *Overall Performance*; WK = *Workload*; NC = *New Colleagues*; CT = *Cooperating Teacher*; SU = *Supervisor*; TM = *Teaching & Managing*; HL = *Helping.*

Structural equation modelling procedures require a sample size large enough to provide reliable parameter estimates (Ullman, 1996). Jöreskog & Sörbom (1993) have suggested a minimum of 5 cases for each estimated parameter in order to obtain stable estimates. To accommodate this guideline, item parcels were constructed to reduce the number of parameters which had to be estimated (West, Finch, & Curran, 1995). An item parcel is the sum or the mean of several items that are assumed to measure the same construct. In our study, we used parcels made up of item means when testing the 4-factor model. Parcelling enabled us to reduce the number of variables in the analysis from 29 to 15. Because the four factors were hypothesised to covary with one another, there were 36 parameters to estimate (15 loadings, 15 error variances, and 6 factor inter-correlations). The sample size for Practicum 1 was 232, which brought the ratio to about 6 cases per estimated parameter.

When testing the 4-factor model, Maximum Likelihood (ML) estimator was employed. The analysis yielded the following values for selected goodness-of-fit indices: RMSEA (*R*oot *M*ean *S*quare *E*rror of *A*pproximation) = .051 (.05 or lower is desired); GFI (*G*oodness of *F*it *I*ndex) = .92; and CFI (*C*omparative *F*it *I*ndex) = .95. All three indices indicate a good degree of fit for the model.

There are some notable similarities and differences in the way students in Singapore and Australia conceptualize their practicum concerns. These differences were identified through separate analyses employed in Singapore and Australia. Singaporean data produced a 7-factor model, identified via exploratory factor analysis. Based on the Australian data, a 4-factor structure was confirmed using structural equation modelling procedures. Since confirmatory factor analysis was not used by D'Rozario and Wong, we were unable to compare the quality of fit of the two models. Assuming that the 7-factor model is a reasonable reflection of the underlying structure inherent in the Singaporean students' responses, the following observations can be made about the two models.

In both countries, students' concerns related to university evaluation of their practicum performance, labelled *University Evaluation* in the Australian model and *Supervisor* in the Singaporean model, were conceptualized identically. This subscale reflects a group of concerns that are particularly robust across dissimilar contexts and cultures. Each of the three remaining Australian subscales is a loose combination of two Singaporean subscales. Thus, in the Australian model: *School Evaluation* combines *New Colleagues* and *Cooperating Teacher*; *Preparation* combines *Overall Performance* and *Workload;* and, *Teaching* combines *Teaching and Managing* and *Helping.* Differences in the structure of these subscales may result from the differences observed in the two countries' cultures and education systems. Observations related to these differences will be drawn out in the discussion.

What Concerns Students Most and Least in the Practicum?

One part of our data analysis was devoted to the item-level comparison between the two countries, in which we compared the frequencies distributions for each of the 29 SPS items. As the Singaporean raw data were not available, the percentages presented in Table 4 of D'Rozario and Wong's (1996) paper were used instead. To make them suitable for our analyses, the percentages were first converted into frequencies. This procedure was carried out on the assumption that in the Singaporean sample, there were no missing data; that is, for each of the 29 items, there were 397 valid responses. These frequencies (excluding the *Not Applicable* category) were then matched with the corresponding Australian data for Practicum 1. Before analysing the combined data set, the two categories indicating high degree of stress (*Most of the Time* and *All the Time*) were combined into a new category, *All/Most of the Time*. The comparisons between the two countries were then carried out on this final set, for each of the 29 items separately.

Table 2 presents percentage distributions for the 29 SPS items, for each country. The percentages show how many students (within each sample) were stressed *never*, *some of the time*, or *most/all of the time*. The corresponding frequencies for each of the 29 items were then submitted to a series of χ^2 analyses to determine the relationship between country of origin and the degree of perceived stress. The result was a significant link between the two dimensions for 25 of the 29 items. The χ^2 (2)'s for these 25 items ranged from 6.40 ($p < .05$) to 175.80 ($p < .001$). There was no statistically significant difference between the two countries for three concerns: High expectations of teaching performance (χ^2 (2) = 0.04, NS); Delivering the lesson (χ^2 (2) = 1.33; NS); and Managing seatwork (χ^2 (2) = 5.62, NS).

Table 2. Percentage Distributions for the 29 Stress Items (by Country)

Item	Country	Never	Some of the Time	Most/All of the Time
Fear of failing the practicum	Sing[a]	34.8	51.5	13.7
	Aust[b]	49.5	43.6	6.9
Striking a balance between the practicum and personal commitments	Sing	36.8	42.5	20.7
	Aust	26.6	49.3	24.0
Having high expectations of my teaching performance	Sing	9.5	50.1	40.4
	Aust	9.4	50.8	39.7
Others expecting me to perform tasks beyond my current competency	Sing	28.3	50.0	21.7
	Aust	33.9	55.3	10.9
Coping with the overall teaching workload (lesson planning, marking)	Sing	5.1	27.6	67.3
	Aust	12.1	60.9	27.0
Managing practicum-related assignments	Sing	16.8	50.0	33.2
	Aust	26.8	53.0	20.2
Communicating with and relating to Principal/Vice-Principal	Sing	64.2	25.7	10.0
	Aust	76.4	19.9	3.7
Communicating with and relating to teachers in the school	Sing	57.0	36.3	6.7
	Aust	74.8	21.6	3.6
Communicating with and relating to my Cooperating teacher(s)	Sing	51.9	34.0	14.0
	Aust	70.5	26.5	3.0
Being observed by my Cooperating teacher(s)	Sing	18.3	53.0	28.7
	Aust	40.3	52.8	6.9
Being evaluated by my Cooperating teacher(s)	Sing	19.0	51.6	29.4
	Aust	36.2	47.9	15.9
Communicating with and relating to my supervisor	Sing	43.2	34.4	22.4
	Aust	60.9	31.5	7.6
Being observed by my supervisor	Sing	9.2	47.6	43.3
	Aust	24.4	51.8	23.8
Being evaluated by my supervisor	Sing	13.1	40.3	46.6
	Aust	30.7	43.5	25.8
Writing detailed lesson plans	Sing	7.6	29.2	63.2
	Aust	31.1	54.1	14.8
Selecting appropriate content for my lessons	Sing	9.9	53.4	36.7
	Aust	18.3	63.1	18.6
Preparing resources for my lessons (e.g., transparencies, worksheets)	Sing	19.8	48.0	32.2
	Aust	33.7	53.6	12.7
Establishing rapport with pupils	Sing	55.7	35.3	9.0
	Aust	65.1	29.9	4.9
Delivering the lesson	Sing	20.9	61.5	17.6
	Aust	20.2	65.1	14.7

Table 2. Percentage Distributions for the 29 Stress Items (by Country) *(continued)*

Item	Country	Never	Some of the Time	Most/All of the Time
Communicating concepts to pupils	Sing	17.8	62.8	19.3
	Aust	28.0	61.2	10.7
Giving appropriate feedback to pupils	Sing	33.5	55.2	11.3
	Aust	47.9	43.6	8.5
Managing groupwork	Sing	16.5	51.7	31.8
	Aust	36.2	52.1	11.7
Managing the individual seatwork	Sing	47.3	45.4	7.3
	Aust	54.7	41.3	4.0
Managing the class and enforcing discipline	Sing	15.4	46.1	38.5
	Aust	14.9	64.4	20.7
Helping pupils with learning difficulties	Sing	17.6	50.7	31.8
	Aust	36.3	53.0	10.7
Helping pupils with emotional/behavioural problems	Sing	23.5	50.1	26.4
	Aust	23.7	58.0	18.3
Teaching mixed ability classes	Sing	14.4	53.4	32.2
	Aust	29.2	56.8	14.0
Marking pupils' written work	Sing	22.0	43.8	34.2
	Aust	60.6	31.5	7.9
Managing time	Sing	14.4	49.6	35.9
	Aust	25.0	52.3	22.7

Note. [a]Singaporean sample N = 340-397; [b]Australian sample: N = 287-309.

A further analysis employing a series of one-sample t tests for each of the 29 SPS items produced an additional significant difference between the two samples for the item Managing seatwork ($M_{Aust} = 1.49$, $M_{Sing} = 1.61$, $t(299) = -3.51$, $p < .001$). Furthermore, as expected, there were 25 items for which Singaporean students reported greater level of concern than Australian students. A notable difference between the groups was the high level of concern indicated by Australian students in striking a balance between the practicum and personal commitments. This was a significantly greater concern for Australian students than for Singaporean students and this was the only item in the survey on which there was significantly greater concern reported by Australian students ($M_{Aust} = 2.05$, $M_{Sing} = 1.91$, $t(303) = 2.79$, $p < .01$).

An inspection of Table 3 reveals that of most concern to Singaporean students were Workload and Lesson Planning, items that were ranked as stressful *most* or *all of the time* by over 60% of this group. Workload was regarded as a stressful activity for the Australian students also, ranked as the second highest concern. For the Australian sample High Expectations of Teaching Performance was the concern held by the highest percentage of students *most/all of the time*. For both groups, Being Observed and Evaluated by their University Supervisor, Managing Time, Managing and Enforcing Discipline, and Managing Practicum-Related Assignments, were all reported as events of concern.

Of least concern to Singaporean students were the following items: Communicating With and Relating to the Principal/Vice-Principal, Communicating With and Relating to Teachers in the School, and Establishing Rapport With Students. Over half the sample reported that these events never stressed them. Other events that generated low levels of concern were Teaching Mixed Ability Classes, Communicating With and Relating to the Cooperating Teacher, and Dealing With

Pupils' Learning Difficulties. Similarly for Australian student teachers, least concern was reported for Relating to the Principal/Vice-Principal, and to Teachers in the School. Among the other events generating low levels of concern were Relating to their Cooperating Teacher, Establishing Rapport With Pupils, and Relating to the Supervisor.

Table 3. Items of Most and Least Concern to Singaporean and Australian Students

	Singaporean Students		Australian Students	
	Of Most Concern:	**%**	**Of Most Concern:**	**%**
1	Coping with the overall teaching workload	67.3	Having high expectations of my teaching performance	39.7
2	Writing detailed lesson plans	63.2	Coping with the overall teaching workload	27.0
3	Being evaluated by my supervisor	46.6	Being evaluated by my supervisor	25.8
4	Being observed by my supervisor	43.3	Striking a balance between the practicum and personal commitments (e.g., family)	24.0
5	Having high expectations of my teaching performance	40.4	Being observed by my supervisor	23.8
6	Managing the class and enforcing discipline	38.5	Managing time	22.7
7	Selecting appropriate content for my lessons	36.7	Managing the class and enforcing discipline	20.7
8	Managing time	35.9	Managing practicum-related assignments	20.2
9	Marking pupils' written work	34.2	Selecting appropriate content for my lessons	18.6
10	Managing practicum-related assignments	33.2	Helping pupils with emotional/behavioural problems	18.3
	Of Least Concern:	**%**	**Of Least Concern:**	**%**
1	Communicating with and relating to Principal/Vice-Principal	64.2	Communicating with and relating to Principal/Vice-Principal	76.4
2	Communicating with and relating to teachers in the school	57.0	Communicating with and relating to teachers in the school	74.8
3	Establishing rapport with pupils	55.7	Communicating with and relating to my Cooperating teacher(s)	70.5
4	Teaching mixed ability classes	53.4	Establishing rapport with pupils	65.1
5	Communicating with and relating to my Cooperating teacher(s)	51.9	Communicating with and relating to my supervisor	60.9
6	Helping pupils with learning difficulties	50.7	Marking pupils' written work	60.6
7	Helping pupils with emotional/behavioural problems	50.1	Managing the individual seatwork	54.7
8	Managing time	49.6	Fear of failing the practicum	49.5
9	Managing the individual seatwork	47.3	Giving appropriate feedback to pupils	47.9
10	Marking pupils' written work	43.8	Being observed by my Cooperating teacher(s)	40.3

At least half the Singaporean students experienced stress at least some of the time for 25 of the 29 practicum-related experiences identified in the survey. For the Australian sample, at least half the students reported being stressed at least some of the time for 22 of the 29 practicum-related experiences.

In broad terms, events involving interpersonal interactions within the school setting were of least concern to both groups of students. The events that generally caused concern were associated with preparation tasks and with being observed and evaluated. The Singaporean and Australian findings support those of Morton et al. (1997) who used Hart's Student Teacher Anxiety Scale (STAS). In their study, Canadian students' anxiety related to being evaluated supports both the Singaporean and the Australian student concerns and, according to Morton et al., is consistent with Hart's finding of student teachers in Great Britain where evaluation anxiety received the highest ratings. (see Morton et al., 1997, p. 70)

Gender Differences

Consistent with the overall finding of significantly greater stress reported by Singaporean students than Australian students, the mean scores on each scale for both males and females were markedly higher for the Singaporean sample than for the Australian sample. Refer to Table 4 for a summary of the comparisons across the SPS subscales.

Table 4. Means and Standard Deviations for the Seven Scales (by Gender and Country)

		Female		Male	
Scale		**Australia**	**Singapore**	**Australia**	**Singapore**
Overall Performance	*M*	10.02	14.59	9.08	14.35
	(*SD*)	(2.31)	(2.95)	(1.87)	(2.80)
Workload	*M*	11.60	20.77	11.16	20.76
	(*SD*)	(3.05)	(3.65)	(2.55)	(4.88)
New Colleagues	*M*	3.99	7.54	3.72	7.04
	(*SD*)	(1.41)	(1.87)	(1.04)	(1.86)
Cooperating Teacher	*M*	3.58	6.37	3.18	6.06
	(*SD*)	(1.23)	(1.54)	(1.26)	(1.63)
Supervisor	*M*	5.69	9.93	5.10	9.72
	(*SD*)	(2.04)	(2.53)	(2.08)	(2.50)
Teaching & Managing	*M*	12.44	20.50	11.56	19.70
	(*SD*)	(2.98)	(3.70)	(2.79)	(4.56)
Helping	*M*	5.67	9.11	5.51	8.56
	(*SD*)	(1.79)	(2.45)	(1.43)	(2.03)

Based on the means for males and females presented for D'Rozario and Wong's seven subscales of the SPS, Singaporean males reported less stress than females on all seven subscales. In the Singaporean study (based on their 7-factor model which did not permit direct comparison with the Australian data), the results showed significantly higher levels of stress for females than for males on two subscales: Overall Performance and Workload ($p < 0.05$). The findings from Singapore related to gender support those of Morton et al. (1997) who found that females reported higher anxiety ratings than males.

In order to compare this result, means were calculated for the Australian sample on the same seven SPS subscales. While Australian males, like their Singaporean counterparts, reported less stress than females, no significant differences were found between the Australian males' and females' levels of concern. This result for Australian students is consistent with analysis of data involving gender on another cohort of students (Murray-Harvey et al., 1999) using the subscales derived from the 4-factor model.

Discussion

How are Singaporean and Australian students' concerns about the practicum conceptualized?

The Survey of Practicum Stresses SPS (D'Rozario & Wong, 1996) was used originally to examine areas of stress experienced by first-year Singaporean teacher education students and was the instrument selected for our own research purposes. A preliminary investigation of the psychometric properties of D'Rozario & Wong's (1996) Survey of Practicum Stresses did not find support in the South Australian data for their 7-factor model. Further analysis found support for a 4-factor model.

We surmise that the different models reflect differences in the way Australian and Singaporean students conceptualize their role as teacher. In Singapore, the emphasis on curriculum prescription, national exams, and large classes suggests an expectation that the teacher's role is to focus student learning on academic achievement; teaching conceptualized quite clearly as instruction. In contrast, beyond developing the intellectual capacities of their students, Australian teachers find themselves accountable for the emotional, physical and mental well-being of their students. Within a very varied Australian education system, teachers also are increasingly expected to be responsive to their school community's expectations. In Australia, it would be difficult to separate out the teaching role from the helping role. This would explain why Australian students conceptualize their teaching role as both teaching and helping and why Singaporean student teachers differentiate between these two roles. This is evident in the way the Australian subscale *Teaching* combines the Singaporean *Helping* items with their *Teaching and Managing* items to form an integrated subscale.

The Australian model also integrates students' relationship concerns with evaluation concerns in the school context as well as in the university context. This integration is consistent with the integration that occurs for both the Australian *University Evaluation* and Singaporean *Supervisor* subscales. These subscales combine the supervisor relationship concerns with supervisor's evaluation concerns. However, the Singaporean model for the school evaluation context, separates out the colleagues' relationships concerns, resulting in two subscales, namely, *Cooperating Teacher* and *New Colleagues.* These two subscales indicate that Singaporean concerns are strongly associated with school evaluation issues as well as relationships with school staff and the principal. On the other hand, Australian students' concerns in the school context are mostly associated with the cooperating teachers' observation and evaluation of them. Teacher and principal relationships were very much lesser concerns for Australian teacher education students. We suggest that a greater relational distance in Singapore between principals and students, teachers and principals, and students and teachers is reflected in the differences between the factors produced in the two models.

Another difference in the underlying structure of Singaporean and Australian student responses worth noting is that of the prevalent concerns of the Australian students in *Preparation.* Compared with the prevalent concerns of the Singaporean students in *Overall Performance* and *Workload*, identified as their current level of competence, preparing resources for lessons and writing detailed lesson plans, Australian students include time management and balance concerns.

The latter refers to balancing the demands on their time of the practicum with the demands of their family or personal lives. For Australian students competence and preparing resources for lessons are secondary concerns to the pressures of workload, detailed lesson plan writing, time and balance demands. Again, we believe that such differences reflect the cultural and education system differences in the two countries, particularly, the strong focus of the Singaporean education system on content, exams, and achieving results.

What concerns Singaporean and Australian teacher education students most and least in their practicum?

Based on item analyses of the percentage of students in the two samples who reported the extent to which they were stressed by a range of practicum experiences, Singaporean students appear to experience significantly higher levels of stress on practicum than Australian students in most areas. However, in both contexts students reported that they were most concerned by much the same events – coping with the workload, high personal expectations of performance, and being observed and evaluated by their supervisor. In all these cases a significantly higher percentage of Singaporean students reported being stressed most/all of the time. Among the items of most concern for Australian students (but not for Singaporean students) was striking a balance between the practicum and personal commitments. In Australia, the students who now comprise the student population bring with them varied life experiences and a range of other competing interests, including work and family responsibilities that need to be balanced with achieving their goal of becoming teachers. The different student profile in the two countries is likely to account for this.

The events that concerned students least were also similar for both student samples. Communicating and Relating to the Principal/Vice-Principal, the student's Cooperating Teacher, Other Teachers in the School, and Establishing Rapport With Pupils, were all items that the highest percentage of Singaporean and Australian students identified as having never stressed them. In summary, preparation and evaluation items generated most stress while interpersonal relationship items were the least stressful in both cultural contexts.

Why do Singaporean students report higher levels of stress?

We suspect that Singaporean students' higher overall levels of concern reflect their more examination-oriented culture. It is also likely that the level of formality in Singapore that exists in relationships between teachers and pupils, and between student teachers and their supervisors, may provide less room for risk-taking and increase performance anxiety. In Australia, conceptualizing helping as part of a teacher's role may actually permit the development of more informal, closer relationships between students and their cooperating teachers (and supervisors), and so reduce the interpersonal concerns that are clearly much greater in Singapore. Similarly, pressures to meet highly structured curriculum expectations and to work within a rigid timetable may partly explain the generally higher levels of concern of student teachers in Singapore. The different cultural contexts may also help to explain differences in the finding of the Singaporean and Australian research between male and female student teachers. D'Rozario and Wong (1996) asked, in relation to their Singaporean students: "Why should female student teachers experience more stress? Could it be that they lack self-confidence? Could it be they have higher personal expectations of themselves? Do they perceive that others have higher expectations of them being 'good' teachers as teaching has traditionally been associated with their gender?" (p. 13). While the same questions may be asked of Australian females, we suggest that the non-significant finding for the Australian sample is a consequence of the impact of policies of affirmative action and antidiscrimination based on gender. The shift towards more gender equality, in the professional arena at least, may help explain the Australian result.

Conclusion

The dearth of literature with a cross-cultural focus is surprising given the increasing mobility of students and teachers between countries, and the growing connections being forged between teacher educators world wide who are developing a shared interest in improving teacher education practices. This study offered the opportunity to contribute to cross-cultural knowledge in teacher education.

References

Bowers, H. C., Eichner, B. K., & Sacks, A. L. (1983). Reducing stress in student teachers. *The Teacher Educator*, 19, 19-24.

Campbell-Evans, G., & Maloney, C. (1995). *Trying to make a difference: Re-thinking the practicum*. Paper presented at the annual conference of the Australian Association for Research in Education, Hobart, Australia, 26-30 November 1995.

Capel, S. A. (1997). Changes in students' anxieties and concerns after their first and second teaching practices. *Educational Research*. 39, 211-228.

D'Rozario, V., & Wong, A. F. L. (1996). *A study of practicum-related stresses in a sample of first year student teachers in Singapore*. Paper presented at the annual conference of the Singapore Educational Research Association and Australian Association for Research in Education, Singapore, 25-29 November 1996.

Elkerton, C. A. (1984). An investigation of stress experienced by beginning teachers during practice teaching. *South African Journal of Education*. 4, 97-102.

Jeans, B. A., & Forth, P. (1995). *The development of professional identity*. Paper presented at the annual conference of the Australian Association for Research in Education, Hobart, Australia, 26-30 November 1995.

Jöreskog, J.D., & Sörbom, D. (1993). *LISREL 8.12a: Users' reference guide*. Chicago, IL: Scientific Software.

MacDonald, C. (1993). Coping with stress during the teaching practicum: The student teacher's perspective. *The Alberta Journal of Educational Research*. 39, 407-418.

Morton, L. L., Vesco, R., Williams, N. H., & Awender, M. A. (1997). Student teacher anxieties related to class management, pedagogy, evaluation, and staff relations. *British Journal of Educational Psychology*. 67, 69-89.

Murray-Harvey, R., Slee, P. T., Lawson, M. J., Silins, H. C., Banfield, G., & Russell, A. (1999). *Under Stress: The concerns and coping strategies of teacher education students.* Third International Conference on Teacher Education, Beit Berl College, Israel. 27 June 1 July.

Ullman, J. B. (1996). Structural equation modeling. In B. G. Tabachnick & L. S. Fidell, *Using multivariate statistics* (3rd ed., pp. 709-811). New York: HarperCollins.

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 56-75). Thousand Oaks, CA: Sage.

Wong, A. F.L., Chiew, G. S., Gopinathan, S., & D'Rozario, V. (June, 1998). Singapore's primary school: social context, teacher preparation and classroom processes involved in high academic achievement. *Education 3 to 13*, 33-13

International Education Journal Vol 1, No 1, 1999
http://iej.cjb.net

An Evaluation of the Implementation of The Dimensions of Learning Program in an Australian Independent Boys School

Murray Thompson
Flinders University *dtmt@senet.com.au*

Key Words: Dimensions of Learning, Rasch Scaling, Hierarchical linear modeling, Australian Schools Science Competition, educational measurement

In order to test the effectiveness of the implementation of the Dimensions of Learning Program in an Australian independent boys school, data from the Australian Schools Science Competition over a number of years from this school were analysed. Subjects were secondary school students from the school over a period from 1994 until 1998 from Grades 8 to 12. The data were Rasch scaled and this allowed the student performance scores for each of the grade levels over the five calendar years of data to be put on one scale. It was thus possible to track the growth of the individual students and the cohorts of students over their time at school. Multi-level analysis using hierarchical linear modelling was employed to test the effects of the hypothesised variables. These variables included grade level, involvement in the Dimensions of Learning Program and IQ. It was found that, in its early stages of implementation, the Dimensions of Learning Program has had a measurable positive effect equivalent to approximately 40 per cent of one year's growth. This result was significant at the 5 per cent level on a one-tailed test. In addition, the Dimensions of Learning Program interacted positively with student IQ, indicating that more able students would appear to profit more from the introduction of the program. This result was also significant at the 5 per cent level. Since the implementation of the Dimensions of Learning Program is a gradual process, these results provide practical evidence of its worth and suggest that as the program becomes more established in a school, further improvements might be evident.

INTRODUCTION

Introduction to Dimensions of Learning

The Dimensions of Learning Program was developed in the United States at McREL, the Mid-Continent Region Educational Laboratory, in Colorado, by Marzano and a team of researchers. The Program brings together what recent educational and psychological research has reported about the way students learn, into an integrated structure, incorporating a wide range of strategies, suitably packaged for use in schools. It grew out of an earlier project, *Dimensions of Thinking* (Marzano et al., 1988). The program is well documented, with a teacher's manual (Marzano et al., 1992a), an assessment manual (Marzano et al., 1993) and a training manual (Marzano et al., 1992b).

Essentially, the Dimensions of Learning Program suggests that for effective learning to take place, the teacher and the learner must attend to thinking and learning in five different areas or dimensions. These five dimensions are encapsulated in the Dimensions of Learning logo, which shows three circles inside a rectangle as shown in Figure 1.

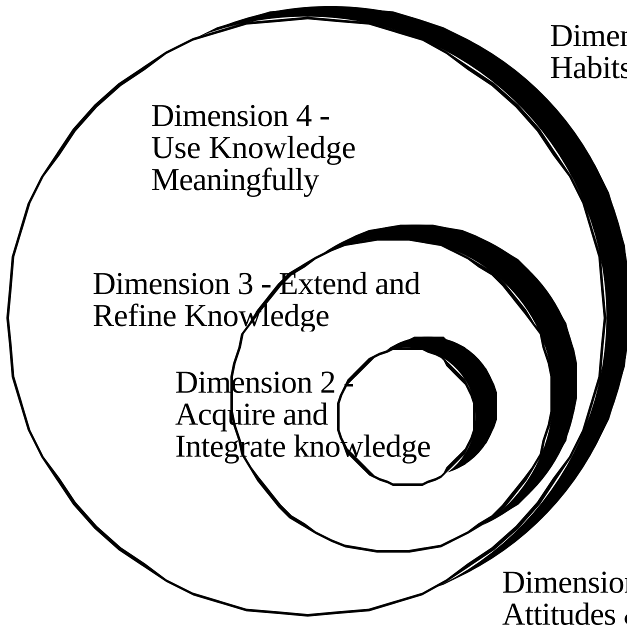


Figure 1. The Dimensions of Learning logo, illustrating the five Dimensions †

† Source: Marzano R. J., (1992)

Dimension 1 is the dimension of ***Positive Attitudes and Perceptions*** which recognises that for students to learn, they must first believe that the learning is within their capability and that the material to be learnt is itself worthwhile. In addition, it indicates that for learning to be effective, students must feel safe, secure and valued in their learning environment.

Dimension 2 is the dimension in which students ***Acquire and Integrate Knowledge.*** It recognises that for learning to be meaningful, the newly acquired knowledge must be built into the already existing knowledge base for each of the students. It breaks knowledge into two types called ***declarative knowledge,*** that is knowledge of facts or knowledge about something and ***procedural knowledge*** or knowledge of how to perform some task. For declarative knowledge, the program provides strategies for constructing meaning, organising the knowledge and for storing the knowledge. On the other hand, for procedural knowledge there are strategies for constructing models of the processes to be learnt, for shaping the skills or processes and for internalising these skills or processes.

Dimension 3 is the dimension in which students ***Extend and Refine Knowledge.*** In this dimension, there are a number of so-called complex reasoning processes which encourage students to examine their knowledge in different ways. This forces them to be engaged in their learning and reinforces their understanding. These complex reasoning processes include comparison, classification, induction, deduction, error analysis, constructing support, abstraction and analysing perspectives. Constructing support involves building evidence to support an opinion or a claim.

Dimension 4 is ***Using Knowledge Meaningfully.*** This dimension uses the complex reasoning processes of decision making, investigation, experimental inquiry, problem solving, invention and systems analysis, which like the Dimension 3 processes, further encourage the students to take apart and re-construct their knowledge.

In Dimension 5, titled ***Productive Habits of Mind***, students are encouraged to develop those mental habits which will enable them to become life-long learners. These mental habits include those of creative thinking, critical thinking and self-regulation. The program suggests a number of characteristics of creative thinkers. It is suggested that creative thinkers engage intensely in tasks even when answers are not readily apparent, push the limits of their knowledge, generate their own standards of evaluation and seek new ways of viewing situations. Critical thinking involves being accurate and seeking accuracy, seeking clarity, restraining impulsivity, taking a position when necessary and being sensitive to the needs of others. Self-regulation involves individuals being aware of their own thinking processes, planning and being aware of the resources necessary to complete a task. In many respects, Dimension 5 of the program sums up the aims of the program.

The implementation of Dimensions of Learning at the school

During 1994, a decision was taken to explore the Dimensions of Learning Program (DOL) to see if it could be implemented at a large Australian independent K-12 boys school. The initial introduction was through the Deputy Headmaster, who attended a training session in New Orleans in 1993. After this training, it was decided by the School Council to institute the program at the school with key objectives being to improve the learning and to provide a theoretical underpinning for learning in the school. As well, it was seen as an important vehicle for the promotion of professional development of the staff. Therefore, a decision was taken to introduce this program, initially with the students in Grade 8, in the first year (1995) and then for students across the school, in subsequent years. It was recognised that such an undertaking was necessarily a long-term one, requiring extensive staff training and a gradual introduction of units of work structured around the five dimensions of learning. Initially, all staff attended an introductory series of seminars in small groups held after school. These seminars were an important part of the professional development program of the school. Naturally, they required a good deal of time and needed staff to be prepared to undertake extra reading. Finding this time in a very busy school program was not easy.

In June 1995, Pollock from the Mid-Continent Region Educational Laboratory, visited the school to promote the program and to give further training. At this time there was some media interest, and the school made a very public commitment to the program. The training sessions consisted of small group seminars, held in faculty groups. With the interest generated by Pollock, staff expanded their planning units of work to grades other than Grade 8 according to the Dimensions of Learning framework. It was very clear in this initial phase, that the implementation of the Dimensions of Learning Program should be a gradual one, in which the learning culture and philosophy of the school would be changed.

There were a number of reasons behind the decision to implement the program at the school. It has strong psychological foundations, deeply rooted in the understandings about how students learn, based on the findings in research in cognitive psychology. It represented a uniform methodology and language of learning which ideally could be consistent in its approach throughout the school and it provided an excellent opportunity to encourage the professional development of every member of staff. Moreover, the comprehensive nature of the program appealed, particularly its recognition of the importance of students as life-long learners. Naturally, it was expected that the program would have the effect of improving the learning of students and that this might be reflected in the external examination results.

The need for evaluation of the program

The decision to take up the Dimensions of Learning Program has not been without cost, both financial and in terms of staff time to re-think the presentation of units of work. In his annual report presented at the end-of-year speech night in 1996, the Headmaster commented on the program;

> *This has had a marked effect on the learning styles of our young men, particularly in the middle years of schooling…Not all students, and indeed not all staff have embraced the ideas with the same degree of clarity and enthusiasm so that…there will be an evaluation undertaken of the achievements within this program. Anecdotal evidence would suggest, however, that many students have gained a great deal from employing the learning techniques suggested within the frameworks of Dimensions of Learning.* (Webber 1996).

There is, therefore, a need to evaluate the effectiveness of the program to see if there has been a change in the learning outcomes of the students involved.

There are a number of underlying difficulties with any project to evaluate the Dimensions of Learning Program. Its introduction, by its very nature, is a long-term process. Staff began by planning a single unit using the Dimensions of Learning Framework and gradually extended their confidence and experience with the program. Thus, there was really no clearly defined beginning to the introduction of the program to the classroom. Rather, it was gradually eased in, initially with only Grade 8, in 1995, and then with the other year levels in subsequent years. Some staff incorporated the ideas eagerly, while others were reluctant to make any change. Thus with such a gradual introduction, there is unlikely to be a dramatic change. Rather, any change might only be evident over a period of several years.

In the United States, action research was conducted to evaluate the Dimensions of Learning program in the Concord-Carlisle School District (Cooper et al., 1996). From the use survey questionnaires and interviews, they reported some positive benefits of the program, including increased learning of course content, more student awareness of thinking processes and enhanced curriculum planning.

A small study relating to an element of the program indicated significant improvements to the long-term learning of Year 11 Physics students using the induction strategy in the program (Thompson 1997). However, what was needed was an evaluation of the success of the overall program. Pressley and McCormick (1995) commented on the earlier Dimensions of Thinking project (Marzano et al., 1992), noting that despite considerable interest in the program, with many attempts to implement its ideas, there had been no research evidence to support the effectiveness of such packaging of a range of strategies. Whilst there had been a great deal of research evidence to support components of the package, it was argued that it did not necessarily follow that the whole package worked.

> *Our greatest concern with this direction is that we have not seen a single convincing evaluation of packages that are comprised of a large number of strategies aimed at diverse cognitive goals.* (Pressley and McCormick 1995, p. 350)

Moreover, as is common with the introduction of a new program, there is no baseline measurement specifically taken for the purposes of comparison, thus making the measurement of any change difficult. In order to evaluate this program effectively there has to be an instrument which is used both before and after the implementation of the program. This instrument must be able to be equated from year to year and between year levels to allow the progress of individual students to be tracked over their time at school for up to five years. As well, it is necessary to track the progress of groups of students, to see if the gradual introduction of the Dimensions of Learning Program has had an effect.

Thus, what is required is an instrument which can be used to evaluate the effectiveness of the Dimensions of Learning Program and a means of equating the results of that instrument both between year levels and from one year to another.

TOWARDS AN EVALUATION STUDY

The Australian Schools Science Competition

The Australian Schools Science Competition is an Australia wide competition which is held every year. The competition is administered by the Educational Testing Centre of the University of New South Wales. Faulkner (1991) outlines the aims of the competition and gives a list of items from previous competitions. Among its aims are the promotion of interest in science and awareness of the relevance of science and related areas to the lives of the students, and the recognition and encouragement of excellence in science. An emphasis is placed on the ability of students to apply the processes and skills of science. Since the science syllabi throughout Australia vary, the questions which are asked are essentially independent of any particular syllabus and are designed to test scientific thinking. Thus, the questions are not designed to test knowledge, but rather test the ability of the candidates to interpret information in scientific and related areas. Thus students may be required to analyse, to measure, to read tables, to interpret graphs, to draw conclusions, to predict, to calculate and to make inferences from data given in each of the questions. Success in this competition requires accuracy and clarity in the thinking processes employed, as well as the ability to make judgements only after careful thought and not from impulse. Thus, in many respects, the Australian Schools Science Competition provides a ready measure of the Dimension 5 processes of creative thinking, critical thinking and self-regulation.

For students in Grades 8 to 12, the test consists of 45 multiple choice questions, each of which has four alternatives from which to choose. It is a timed test and the participants are given one hour to complete the questions. The students are required to record their answers by filling in a bubble on the prepared answer sheet for the chosen alternative in each question. The answer sheet is then marked using an optical scanner. There is considerable overlap in the questions from one grade level to another, with a significant number of items administered to more than one grade level. Indeed, a few items in each grade are common to more than two of the grade levels. However, there is no attempt to repeat items from previous years. Data of the results obtained from students for 1994, 1995, 1996, 1997 and 1998 were available.

It was necessary to establish a means of comparing the results of each of the year groups and comparing the results from one year to another, in order to track the progress of both individuals and groups of students.

Rasch Scaling and Item Response Theory

The Rasch model of test scaling has come into prominence over recent times and it has become the basis of many testing programs throughout the world. Snyder and Sheehan (1992) provide a good introduction to the principles of Rasch scaling. The power of the Rasch scaling method is that the measurement on the scale of the performance abilities of the students taking a test is independent of the test items and that the difficulty level on the scale of the test items is independent of the group of students used to calibrate the scale and the test items. Essentially, this model assumes that the likelihood of a student correctly answering a question depends upon the difference between the difficulty of the item and the performance level of the student, both measured along a continuum, known as the latent trait. This model is based upon a number of assumptions and requirements. The first of these is the so-called "know correct" assumption which simply suggests that if a student knows the correct answer, then the student will probably get the question correct. A second requirement is that of unidimensionality so that the test must only be measuring one underlying trait or such traits working in unison. Weiss and Yoes (1991) make the point that this type of test is not valid under speeded conditions. A third requirement which is considered is that the questions should be not related to the extent that answers given to one question should not affect those in any other question. Thus, it is assumed that there should be no cues given in any of the questions which may help in other questions. Likewise, each student responding should provide answers that are independent of the answers given by other students.

A number of computer programs have been developed to assist with the analysis of data from test items. One of these is the Quest program, developed by Adams and Siek-Toon Khoo (1993), which allows the results of tests to be analysed to determine whether they fit the Rasch model. It provides estimates, both of the abilities of the students and the item difficulties of the test. In a recent study (Thompson 1998), it was found that the Australian Schools Science Competition data fit the Rasch model well, in spite of the timed nature of the test, allowing the estimates of the item difficulties and the student abilities to be plotted on the same scale. Moreover, the inclusion of common items in the tests for the various grade levels allows the data for different grade level groups to be placed on the same scale, providing facility for the comparison of the different grades. It was found that the most suitable method for this equating process was the concurrent equating method. This involves scoring all of the items and subjects at the one time, relying on the common items to establish the difficulty levels of all of the items across the range. This method has been tested in comparison to other equating procedures by Mohandas (1998) and provided good agreement with the expected results.

It was also necessary to establish a means of relating the results of the test of each calendar year to those of other years, so that the progress of individual students could be tracked, by using a scale common to all years. In order to overcome this problem, students in Grades 8, 9 and 10 were given four items appropriate to their grade levels from the 1993, 1994, 1995, 1996 and 1997 tests as a practice a few days before the 1998 test. These items formed an equating test and were then included with the 1998 test to establish item difficulties for the items in the previous years, alongside the 1998 scores. These difficulties, thus calculated, were then used as anchors in the anchoring method to determine the difficulties of the items and the student abilities in the previous years.

It was evident that the Australian Science Competition could be used to provide data for analysis which would enable comparison between groups of students and the progress of groups and individuals to be tracked. Thus the data for students in each of Grades 8, 9, 10, 11 and 12 from the Science Competitions of 1994-1998 could be put on one scale so that comparisons could be made and some measurement of growth could be established. Moreover, with the capacity to equate the scores between years, it was possible to make an estimate of the magnitude of change per year over time.

In this way the data from the Australian Schools Science Competition, when Rasch scaled, could be used to make comparisons between groups of students and to measure the growth shown by groups and individual students over several years and thus to provide some evidence for the evaluation of the Dimensions of Learning Program. It must be recognised, however, that there may be a range of factors which could be responsible for any change in the performance of the students, as measured by the scaled scores of the Australian Schools Science Competition. In the absence of any possibility of providing control through an experimental design, statistical control, using regression analysis procedures had to be employed to allow for the effect of such factors.

Hierarchical Linear Models

Factors affecting the results of the science competition

In the previous sections it has been argued that the Australian Schools Science Competition could provide a measure of the effectiveness of the Dimensions of Learning Program by reflecting some of the characteristics of Dimension 5, such as critical thinking, creative thinking and self-regulation. It must be admitted, however, that there would be many reasons why an individual student's performance could show change over time. First, on each successive occasion, each student is a year older and change could be expected to come from the extra year of maturity and educational experience, quite independent of the intervention of the Dimensions of Learning Program. Other factors which could be expected to affect student performance include the innate ability of the student, as might be measured with a standardised IQ test, as well as the effect of the grouping of students, or the way in which the students interact with one another. Thus, it could be argued that the way in which students cooperate with one another and support one another in their learning could have a significant effect on the performance of the group. In order to measure the effect of the Dimensions of Learning Program on the performance of the students, the individual effects of factors, such as those described above, must be controlled in statistical analysis.

The analysis of change

Willett (1997) has argued that in order to measure change, there must be a number of points at which data are recorded. It is suggested that the more measures of a particular variable taken over the course of an experiment, the more likely it is that the errors are reduced. It is further suggested that the analysis of the data using a hierarchical linear model can test for the operation of some underlying causes of the change. These hierarchical linear models are discussed by Raudenbush and Bryk (1997), Bryk and Raudenbush (1992), and Keeves and Sellin (1997). Such models allow a researcher to postulate and subsequently to test statistical hypotheses associated with relationships between the outcome variable and the factors which may affect it. In hierarchical linear models, the researcher can examine the effect of the various factors, both within and between individuals and at the group level and any possible interactions between them. The outcome variable is represented as a function of the various characteristics.

Such hierarchical linear models involve an examination of within student performance which changes over time and how this is affected by characteristics of the students. It is thus often referred to as a two-level model, with one level representing the individual variations within student and the second level which reflects the variations between students. In this study, it is necessary to analyse performance as measured on the science test across several years of schooling. The first level is then the performance within students and how it is distributed over time. The performance of an individual in the test might be represented at Level 1, the micro-level, or within student level by :

$$Y_{ij} = \beta_{j0} + \beta_{j\,1}\,(\textit{grade level}_{ij}) + \beta_{j\,2}(\textit{exposure to DOL}_{j2}) + r_{ij}$$

In this equation Y_{ij} represents the performance of student j at grade level i and β_{j0} represents the baseline performance of student j. Each of the coefficients represents the extent to which the performance of a student is affected by the parameter in the bracket. Thus coefficient β_{j1} represents the effect of the grade level and its parameter value in the bracket represents the particular grade level. The coefficient β_{j2} represents the effect of the Dimensions of Learning Program, for those exposed to the program, and its parameter indicates whether or not student *j* was exposed to the Dimensions of Learning program in that particular grade at level *i*. The term r_{ij} represents the random error. An important feature of hierarchical linear models is that these coefficients will vary from student to student.

At the second or macro level of a hierarchical linear model, each of the coefficients in the Level 1 equation is expressed as a linear equation of Level 2 variables at the second or between student level. For example, the coefficient β_{j1}, the effect of grade level on student performance of student j may be expressed as a function of a range of parameters at the student level. For example, a researcher may model this as follows.

$$\beta_{j1} = \gamma_{01} + \gamma_{11}\,(\textit{cohort1993}) + \gamma_{21}\,(\textit{DOL1}\,) + \gamma_{31}\,(\textit{IQ}) + u_{j1}$$

where γ_{01} represents the average level across the students of the coefficient β_{j1}, γ_{11} represents the influence of being in the 1993 cohort of students, γ_{21} represents the influence of the Dimensions of Learning Program on growth, and γ_{31} represents the influence of student IQ as determined across the range of students. The term u_{j1} represents the error term associated with the student, unexplained by the model. Thus, in a sense, the researcher is modelling whether the characteristics of a student can be represented by a linear combination of the various parameters and in turn, whether these characteristics influence the individual student performance.

It can be seen then that there is a layered or hierarchical model being used. The values of the various coefficients need to be estimated using the available data from each of the years for which data from the Australian Schools Science Competition are available. Recent advances in computational technology make such estimations possible. One program which does this by a iterative method using empirical Bayes estimation procedures based on the maximum likelihood estimates is HLM developed by Raudenbush and Bryk (1996). With this facility, it is possible to estimate the effect of the various parameters and their inter-relationships at each of the levels of the hierarchical linear model.

It follows then, that it may be possible using a hierarchical linear model, to partition off the effects of the variables such as year level, IQ and the effects of student grouping mentioned above as well as the effect of the Dimensions of Learning Program which is of interest in this study and to estimate the effect of each on student performance.

RESEARCH QUESTIONS

It has been suggested that the implementation of the Dimensions of Learning Program needed to be evaluated. The difficulties of making such an evaluation, at this school, have been discussed. The long-term and gradual nature of the introduction of the program, the lack of specifically designed base-line data for the purpose and the varying degree of staff enthusiasm for the program are among the problems involved.

The questions to be addressed in this study are:

1. Is student learning more effective using the Dimensions of Learning Program?
2. Can the influence of the Dimensions of Learning Program on student performance be estimated?
3. Does the influence of the Dimensions of Learning Program interact with grade level?
4. Does the influence of the Dimensions of Learning Program interact with the IQ of the students?

The answers to these questions are important because they may provide effective ways to improve the education of young people. The answers may provide concrete evidence of a positive effect of the complete Dimensions of Learning package and the opportunity of sharing the strategies with the wider educational community.

THE RESEARCH METHODS

The Data and their Sources

The Science Competition data

From 1993 until 1998, students from the school sat for the Australian Schools Science Competition. The results from each of these years were sent to the school on paper. Over the these years, almost all of the students in Grades 8 and 9 have done this competition along with most of the students in Grade 10. In Grades 11 and 12, students have had the choice to take part in the competition and generally the more able students have chosen to do so.

Unfortunately, the 1993 data were not available in sufficient detail. Since 1994, the competition organisers have sent a comprehensive report to the participating schools. As well as the overall result for each student, there has been a record of the answers given to each question by each student. This has been provided as a simple text list, such as is shown below.

```
STUDENT MARK BCBDAAABDCADDBCBBABADDBCCCBCDBDADDABCBCABDBDB
```

These lists were scanned using an optical scanner, converted to a word processing document, checked for correctness, and arranged into appropriate columns. This was then exported into a text file and then into a spread-sheet. The papers were carefully checked to find the common items across the tests for each grade. The number of different items in each year varied considerably, with 106 different questions in 1995 and 160 in 1996. Prior to 1996, the students in Grades 11 and 12 sat for the same paper. In each year of the competition, there was considerable overlap in the items across the grade levels, as indicated earlier, with quite a large proportion given to more than one grade level. The items were arranged into a grand order, starting with Grade 8 items, then those questions given to Grade 9, but not Grade 8 and then those given to Grade 10 and neither Grade 8 or 9 and so on. Once this order was established, each of the columns of item responses had to be transported to the column corresponding to its grand item number. This resulted in a spread sheet of about 180 columns and about 400 rows, the actual size depending on the number of common items for each year and the number of subjects taking the competition each year. One particular difficulty encountered was the changing of the order of the alternative responses from one grade level to another, necessitating some complicated re-coding of responses.

Once the spread-sheet manipulations were complete, they were saved as a text file and then opened as a word processing document. Some manipulation in this format was necessary to ensure that each student ID (identity) and the sets of responses were in the correct columns. Also at this stage, each missing response was re-coded to a dummy response E. This of course meant that missing responses were recorded as incorrect. The purpose of this was to distinguish the missing responses from the nil responses to those questions which were included in the analysis, but which not all students were required to do. Finally the array of letters was saved as a text file ready to be used as input into the Quest program used to calibrate and score the test data.

The Quest analysis was run for each of the years, with the extra questions from the equating test included at 1998 and then the item difficulties, as determined from the combination of the equating test and the 1998 test, were used to anchor the tests from 1994 - 1997. After running each analysis, the results were scanned to ensure that the questions fitted the Rasch model, with any question not doing so being rejected from the analysis. The output of the these analyses provided an estimate of performance abilities from each of the years 1994 through to 1998 on a scale that measured across years and across grade level. These student performance data were then available for further analysis.

The measurement of the IQ data

Early during Grade 8, the first year of secondary schooling, all students at this school took the ACER Intermediate Test G. This is a test of general reasoning ability which comprises items to assess both verbal and numerical ability which relate generally to intelligence and learning ability in schools. The purpose for this testing is to provide information which helps teachers to assess the students' abilities and to inform advice given to students about future courses and careers. The manual by de Lemos (1982) for the administration and interpretation of the test provides further details. These IQ data for most of the students were then available for inclusion in this study. Those for whom no data were available joined the school at some later time or were overseas students with a non-English speaking back-ground.

The Dimensions of Learning variables

The Dimensions of Learning Program was first introduced to students in 1995. During this year, programs of work using the Dimensions of Learning framework were presented to Grade 8 students. In the following year, the use of the program was extended to all years in the school. Thus, there were varying degrees of the use of the program, with it being used for some of the cohorts from the beginning of their secondary schooling in Grade 8 and others being introduced at a later stage. In order to cope with this complication, a dichotomous variable, DOL1, was introduced at Level 2 to indicate involvement with the Dimensions of Learning Program. In the first instance, students who experienced the program from the beginning of their secondary schooling in Grade 8 were assigned the value 1, whilst those who either did not experience the program, or were exposed to it at a later stage were assigned the value 0. This variable was called DOL1. As well, if in subsequent years, students were exposed to the program, a second dichotomous variable, called DOL, linked to the years in which the student was exposed to the program was introduced as a Level 1 variable. Thus, a student who commenced secondary schooling in 1994 would receive a "0" for the DOL1 at Level 2. At Level 1, this same student would receive, for the DOL variable a "0" in 1994 and 1995 but in 1996, 1997 and 1998, would receive a "1", denoting involvement with the program, during those three years. On the other hand, a student who commenced in 1995 in Grade 8 would receive a "1" for this variable throughout this study for both the Level 1 DOL variable and the Level 2 DOL1 variable. A third variable was introduced, DOL2, for an identified group who, during their Grade 8, in 1995, had a larger input of Dimensions of Learning than the other groups. Table 1 shows the various cohorts and their treatments. The cohort groups are labelled according to the year in which they were in Grade 8. It should be noted that the school year in Australia is from January until December.

Table 1. The various student cohorts and their involvement with the Dimensions of Learning program over the years.

	1994	**1995**	**1996**	**1997**	**1998**	**Number in cohort group**
Cohort group 1993	Grade 9 no DOL	Grade 10 no DOL	Grade 11 DOL	Grade 12 DOL	-	51
Cohort group 1994	Grade 8 no DOL	Grade 9 no DOL	Grade 10 DOL	Grade 11 DOL	Grade 12 DOL	103
Cohort group 1995	-	Grade 8 DOL	Grade 9 DOL	Grade 10 DOL	Grade 11 DOL	98
Cohort group 1996	-	-	Grade 8 DOL	Grade 9 DOL	Grade 10 DOL	103
Cohort group 1997	-	-	-	Grade 8 DOL	Grade 9 DOL	97

Student cohort variable

It was recognised that the grouping of students may well have a significant effect on the learning outcomes of the students. In order to isolate the effect of the Dimensions of Learning Program, it is important to recognise the importance of the interrelations between the students with one another and their learning. The Rasch analysis of the abilities and the item difficulties included data from all of the students who sat for the tests. However, the subsequent analysis included only those students who had data recorded on at least two data points. Thus, only students who sat for at least two Australian Schools Science Competitions from 1994 until 1998 were included in the subsequent analysis. In addition, six more dichotomous variables were introduced, at both Level 1 and Level 2, to reflect the cohort grouping of the student.

THE RESULTS

A number of models were explored but the one which best explained the data is given below.

Level-1 Model

$$Y = B_0 + B_1*(LEVEL) + B_2*(DOL) + R$$

In this Level 1 model, the outcome variable Y, the Rasch scaled performance scores measured by the Science Competition test are equal to an intercept or base level B_0, plus a growth term due to the increase in grade and educational experience as each year passes, with associated slope B_1. As well, there is a growth and associated slope B_2, due to involvement with the Dimensions of Learning Program plus an error term, R.

In the Level 2 model, the effect of the Level 2 variables on each of the B terms in the Level 1 model is given.

Level-2 Model

$$B_0 = G_{00} + G_{01}*(93) + G_{02}*(96) + G_{03}*(IQ) + U_0$$

$$B_1 = G_{10} + G_{11}*(DOL1) + U_1$$

$$B_2 = G_{20} + G_{21}*(IQ) + U_2$$

Values of each of these terms are estimated and the level of statistical significance evaluated to assess the effect of each of the terms.

Initially, the HLM program makes estimates of the various values of the slopes and intercepts and then using an iterative process improves the estimation using a maximum likelihood estimation and empirical Bayes procedure.

Table 2 shows the final estimation of the fixed effects of the model, while Table 3 shows the final estimation of the variance components.

In order to calculate the amount of variance explained by the model, a null model, with no predictor variables was formulated. The estimates of the variance components for the null model are shown in Table 4.

Using the data from Tables 3 and 4, the amount of variance explained is calculated as follows:

$$\text{Variance explained at Level 2} = \frac{0.543 - 0.0854}{0.543} = 0.843$$

$$\text{Variance explained at Level 1} = \frac{0.329 - 0.227}{0.329} = 0.314$$

As well, ρ the intraclass correlation can be calculated.

$$\rho = \frac{\tau_{00}}{\tau_{00} + \sigma^2} = \frac{0.543}{0.543 + 0.329} = 0.623$$

Table 2. The final estimation of the fixed effects

Fixed Effect	Coefficient	Standard Error	T-ratio	P-value
For INTRCPT1,B0				
INTRCPT2,G00	0.370	0.049	7.6	0.000
93,G01	0.333	0.084	4.0	0.000
96,G02	-0.149	0.068	-2.2	0.028
IQ,G03	0.028	0.002	11.8	0.000
For GRADE LEVEL slope, B1				
INTRCPT2,G10	0.263	0.031	8.6	0.000
DOL1, G11	-0.018	0.037	-0.5	0.624
For DOL slope, B2				
INTRCPT2,G20	0.106	0.059	1.8	0.036 [a]
IQ, G21	0.0055	0.003	2.0	0.043

[a] a one-tailed test of significance is employed as hypothesised

Table 3. Final estimation of variance components

Random Effect	Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1, U0	0.292	0.085	95	115.7	0.074
GRADE LEVEL slope,U1	0.025	0.001	97	87.6	>.500
DOL slope, U2	0.220	0.048	97	77.3	>.500
level-1, R	0.475	0.227			

Table 4. Estimated variance components for the null model

Random Effect	Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1, U0	0.737	0.543	451	2528	0.000
level-1, R	0.574	0.329			

This intraclass correlation represents the variance within students compared to the total variance between and within students. Thus 62 per cent of the variance between and within students can be explained as being within the students, whilst the remaining 38 per cent of the variance is between students. This is not surprising, since it suggests that most of variance is as a result of the development of the students over time.

Discussion and interpretation of the results

In order to interpret the results Table 2 is examined. The term G_{00} represents the grand mean of the ability variable, that is the average of the science competition achievement scores, after Rasch scaling, across all grade levels and across all years. Thus the average score is 0.370. The coefficient G_{01} represents the conditional main effect of the 1993 cohort of students. Since this value is positive and since it is statistically significant it can be concluded that, taking into account IQ scores and the effect of the Dimensions of Learning variable, the 1993 cohort of students are exceedingly capable. On average, the student from this cohort performed 0.333 times better than the average. Conversely, the coefficient for the 1996 cohort, G_{02}, is lower than the average by -0.149. This is also statistically significant and again takes into account the variation brought about by the other variables such as IQ and involvement in the Dimensions of Learning Program. Anecdotal evidence suggests that the 1993 cohort was a particularly scholarly group who worked very well together, supporting the learning of each other. Indeed, the inter-relations between this group were especially positive and it is not surprising that the results in the science competition reflect this effect. On the other hand, it has certainly been the experience of the staff of the school that the 1996 group of students are less cooperative with one another. This may explain the differences in these cohorts from the others, but further investigation is necessary to support this view. The other cohorts, 1994, 1995 and 1997 proved not to vary significantly from the grand mean. In other words, membership of these cohorts of students did not appear to be significantly associated the individual student performance. The value G_{03} represents the effect of the IQ of the students. Clearly this has a significant effect and even though the value seems very small, 0.028, it must be remembered that it involves a metric coefficient for a variable whose mean value is in excess of 100 and has a range of over 50 units.

The next important value is B_1, the grade level slope, which reflects the contribution to the growth of the student performance, on average, due to their increase in age and educational experience. It indicates that on average, the estimated performance on the science competition tests goes up by 0.263 every year, after controlling for the effects of the other variables. This also is statistically significant. The Level 2 variable which may have affected this slope was the DOL1 variable, which indicated an involvement with the Dimensions of Learning program from entry to Grade 8. It can be seen from the table that this effect is not statistically significant. Likewise, the effect of the extra involvement in the Dimensions of Learning program is not statistically significant and has been removed from the model.

Of particular interest to this study is the coefficient B_2, the contribution to the performance variable from the involvement in the Dimensions of Learning Program. As Table 6 indicates, involvement in the Dimensions of Learning Program contributes 0.106 to the performance of the students. Since the Dimensions of Learning variable, DOL, is a dichotomous one, this will change from zero to one at most once only during the five years. Comparing this to the 0.263 value for yearly growth it is seen that the Dimensions of Learning Program has added the equivalent of approximately 40 per cent of one year's growth to the performance of the students. On a one-tailed test, this result is significant at the five per cent level, providing good support for the program. In addition, the positive value of G_{21}, 0.0055, indicates that the Dimensions of Learning Program is more effective with those students of higher ability, with this result being statistically significant at the five per cent level.

CONCLUSION

The purpose of this evaluation study is to explore the effect of the implementation of the Dimensions of Learning program in one school. The difficulty of having no specific baseline data to measure the effect is overcome by the use of the already existing Australian Schools Science Competition data. Further, the problems associated with the gradual introduction of the program are overcome through the use of the HLM methodology which has allowed for the progressive inclusion of the various cohorts of students as they became involved in the program.

The lack of available data for some students has certainly affected the reliability of certain relationships in the estimation and it is clear that it is important to continue to expand this research since the 1996, 1997, 1998 cohort data become more complete as the students sit for more Science Competitions and complete their schooling. Similarly, if the more detailed 1993 Science Competition data becomes available in the future, its inclusion would certainly improve the strength of the findings.

It is also important to recognise that the implementation of the Dimensions of Learning Program at this school is still only in its early phase. As the learning culture in the school changes, as teachers become more familiar with the methodology and the terminology and as students themselves recognise the importance of utilising some of the learning strategies to which they are being exposed and apply them in their own study and learning, it might be anticipated that further positive growth is likely. Thus, the positive result at such an early stage of the implementation of the program is most encouraging.

ACKNOWLEDGMENTS

This piece of work would not have been possible without the patience and understanding of a number of people. I am indebted to Professor John Keeves for his insight into the complex problems of educational research and his encouragement to keep persevering, even when the answers were not readily apparent. My thanks too to Mr Milton Haseloff, former Deputy Headmaster of the school, whose energy and vision ensured that, despite the difficulties, the Dimensions of Learning Program was implemented and whose continued support has been much appreciated. Dr Brian Webber, Headmaster of the school, has continued to be most supportive and interested in this project. I am indebted, too, to my colleague Mr David Weise who first alerted me to the significance of the Australian Schools Science Competition as an important instrument to measure student thinking in science and to Mr Ken Watson who provided valuable criticism of this paper. My thanks too to all of my colleagues in the Science Faculty who helped to administer the equating tests. Dr John Faulkner, of the Educational Testing Centre of the University of NSW, provided some useful information about the Science Competition.

REFERENCES

Adams, R. J. & Khoo, Siek-Toon (1993) *Quest the interactive test analysis system* Hawthorn Vic: ACER.

Bryk, A. S. & Raudenbush, S.W. (1992) *Hierarchical linear models: applications and data analysis methods* Beverly Hills, Ca: Sage

Bryk, A.S., Raudenbush, S. W., & Congdon, R.T. (1996) *HLM for Windows version 4.01.01* Chicago: Scientific Software

Cooper, L. A., Devlin, H., Leone, J., MacLean, D., Pattie, J., Penniston, L., Moser, P. (1996). Report to the Concord and Concord-Carlisle Committees on the Dimensions of Learning Action Research Project. Unpublished report

de Lemos, M. M. (1982) *ACER Intermediate test G Manual for administration and interpretation* Hawthorn, Vic : ACER

Faulkner, John (ed) (1991). *The Best of the Australian Schools Science Competition* Rozelle, NSW : Science Teachers' Association of New South Wales

Keeves, J. P. and Sellin, N. (1997) Multilevel analysis. In J. P. Keeves, (ed) *Educational research, methodology and measurement* (2nd ed.), Oxford: Pergamon, pp. 3978-3987

Marzano R. J., (1992) *A different kind of classroom : Teaching with Dimensions of Learning* Alexandria Va.: Association for Supervision and Curriculum Development.

Marzano, R.. J., Brandt, R. S., Hughes, C. S., Jones, B. F., Presseisen, B. Z., Rankine, S. C., and Suhor, C. (1988) *Dimensions of Thinking: A framework for curriculum and instruction* Alexandria Va.: Association for Supervision and Curriculum Development.

Marzano, R. J., Pickering, D. J., Arrendo, D. E., Blackburn, G. J. Brandt, R. S. and Moffett, C. A. (1992a) *Dimensions of Learning teacher's manual* Alexandria Va.: Association for Supervision and Curriculum Development.

Marzano, R. J., Pickering, D. J., Arrendo, D. E., Blackburn, G. J. Brandt, R. S. and Moffett, C. A. (1992b) *Dimensions of Learning trainer's manual* Alexandria Va.: Association for Supervision and Curriculum Development.

Marzano, R. J., Pickering, D.J. and McTighe J. (1993) *Assessing student outcomes: performance assessment using the Dimensions of Learning model* Alexandria Va.: Association for Supervision and Curriculum Development.

Mohandas, R. (1998) Test equating, problems and solutions: Equating English test forms for Indonesian junior secondary school final examinations administered in 1994. Unpublished Masters Thesis, The Flinders University of South Australia.

Pressley, M. and McCormick, C.B. (1995) *Advanced educational psychology for educators, researchers, and policymakers.* New York : Harper Collins

Raudenbush, S.W. & Bryk, A. S. (1997) Hierarchical linear models. In J. P. Keeves, (ed) *Educational research, methodology and measurement* (2nd ed.), Oxford: Pergamon, pp. 2590-2596

Raudenbush, S.W. & Bryk, A. S. (1996) *HLM Hierarchical linear and nonlinear modeling with HLM/2L and HLM/3L programs* Chicago: Scientific Software

Snyder, S. & Sheehan, R. (1992) Research methods The Rasch measurement model: An introduction. *Journal of Early Intervention* 16 (1), 87-95

Thompson, M. J. (1997) Induction, a powerful strategy to extend and refine knowledge. Unpublished paper, The Flinders University of South Australia

Thompson, M. J. (1998) The Australian Schools Science Competition -

A Rasch analysis of recent data. Unpublished paper, The Flinders University of South Australia

Webber, B. J. (1996) Report to the school speech night for 1996

Weiss, D. J. and Yoes, M. E. (1991) Item response theory. In Hambleton, R. K. & Zaal, J. N. (eds) *Advances in educational and psychological testing* Boston : Kluwer, pp. 69-95

Willett, J. B.(1997) Measurement of change. In J. P. Keeves, (ed) *Educational research, methodology and measurement* (2nd ed.), Oxford

Learning Styles and Perceptions of Self

Ritu Dangwal
Centre for Research on Cognitive Systems, India *ritud@niit.com*

Sugata Mitra
Centre for Research on Cognitive Systems, India

Kolb's Learning Styles Inventory is one of the most commonly used instruments to identify a learner's style. This paper attempts to find out whether our own perception of our learning style matches other people's perception of our learning style. We have used a 360-degree technique to measure an individual's learning style as perceived by others. A Mean Learning Style and a Learning Style Vector are defined to represent a collective perception of an individual learning style. The results show consistent differences between our own and others' perception of our learning styles. These differences are analysed and validated using a placebo method and found to be accurate descriptions of perception. The results suggest that a learning styles inventory test, used in this manner, could be a measure of public and private perceptions of self.

Key words: vector, perception, abstract conceptualisation, concrete experience, placebo.

Introduction

Learners learn in different ways. Several studies discuss learning styles and find that learning depends upon many personal factors and everyone has a distinct learning style (*S. Montgomery, 1996; Mumford & Honey,* 1996).

In the present study, results reveal that there is a discrepancy in our perception of our learning style and in others' perception of our learning style.

The term "Learning Style" has been defined as "the composite of characteristic cognitive, affective, and physiological factors that serve as relatively stable indicators of how a learner perceives, interacts with, and responds to the learning environment" (*Keefe, 1979*). According to *Cronbach & Snow* (1977), learning styles could be used to predict what kind of instructional strategies or methods would be most effective for a given individual and learning task.

One way to optimise learning could be to find out a person's characteristic style of learning, if any, and then match the learning environment to the person.

The Learning Style Inventory (LSI) is a test that assesses the characteristic learning style of a person.

Presently, there are different learning styles models, which are based on different psychological theories. Models based on Personality include *Witkin's* (1954) and *Myers-Briggs Type Indicator* (Myers,1978). *Schmeck's (*1983) and *Kolb's* (1984) models are examples of the Information Processing approach. *Reichmann and Grasha* (1974) models are based on Social Interaction. And finally, *Keefe* (1989) and *Dun & Dun* (1978) models are based on the multidimensional factors within Human Information Processing. All these models stress the importance of identifying and addressing individual differences in the learning process.

The most commonly used inventory is the one proposed by *Kolb* (1984) which is based on Piaget's model on L*earning & Cognitive Development* and Jung's personality type, i.e. *Myers-Briggs Type Indicator* (MBTI).

We administered Kolb's Inventory to a large number of persons in India and were faced with difficulties such as:

1. The rationale for the statistical treatment used to derive the results.
2. The rationale behind choosing buffer items.
3. Lack of clarity in the meaning of the adjectives used in the inventory.

In order to address the Indian context, the inventory had to be adapted to suit the Indian population. We adapted Kolb's Learning Styles (LSI) Inventory to the Indian context (Dangwal and Mitra, 1997). As in the case of Kolb's original inventory, the present inventory involves four principal stages**:** Concrete Experience (CE): learners are able to involve themselves fully, openly, and without bias in new experiences; Reflective Observation (RO): learners are able to reflect on and observe their experiences from many perspectives; Abstract Conceptualization (AC): learners are able to create concepts that integrate their observations into logically sound theories; Active Experimentation (AE): learners are able to use these theories to make decisions and solve problems. The CE/AC and AE/RO dimensions are polar opposites as far as learning styles are concerned and it identifies four types of learners depending upon their position on these two dimensions. The four types of learner are: Analytical, Imaginative, Precision and Dynamic.

Further, learning activities can be divided into two categories: Perception and Processing. Perception is measured by the CE/AC continuum. For example, some people best perceive information using concrete experiences (like feeling, touching, seeing and hearing) while others best perceive information abstractly (using mental of visual conceptualisation). Once information is perceived it must be processed which is measured by the AE/RO continuum. For example, some people process information best by active experimentation (doing something with the information), while others process best by reflective observation (thinking about it).

Use of a 360 degree technique

Is the Learning Styles test telling us what our learning style is or is it telling us what we would like our learning style to be? We felt that if the learning style identified by the test for an individual could be corroborated with the perceptions of that individual's learning style by others who know him or her, then there would be reason to believe that the test is measuring what it is supposed to measure. We could find no test that allows us to measure how others' perceive our learning style. We, therefore, adopted a 360-degree technique. In this technique, we identify a person X and ask other people who know X to take the learning styles test but mark each item as they think X would have marked them. In the present experiment, we have compared self perception of learning style with others' perception of an individual's learning style using the learning styles inventory (Dangwal and Mitra, 1997).

Purpose of this study

We felt that the following questions are important if the results of any learning styles test are to be useful. These questions may be relevant for any test but in the present paper we have taken the context of the LSI.

1) Do we identify our learning style as it is or as we would like it to be?

   This is important in order to determine what style would best help a learner achieve a learning objective. If a LSI test merely identifies a learners desired style as opposed to how he or she actually learns, the test would have limited usefulness for designing instruction for that learner.

2) How do other people perceive our learning style?

   This is important if teachers, or teaching systems, are to use LSI test results to modify their style of instruction. If others perceive our learning styles differently from how we perceive it, the difference would need to be understood.

3) Does the LSI correctly reflect our perception of our own as well as of others learning styles?

   It is necessary to know whether the description of a learning style for a particular person as given by a LSI test, is consistent with that person's own as well as other's experiential assessment of his/her learning methods. If the experiential conclusion differs from the test results, the differences would need to be understood.

Methodology

We decided to use a "360 degree" method to answer some of the questions above. A person (say P) would take the LSI test as usual. Others taking the LSI test would follow this, but they would mark the items in the test as they think P would have marked them. The results of the latter tests would constitute other's perception of P's learning style. The selection of the group is very crucial. The target group must know the person P "well enough" to take the test as P.

We report the study in two parts:

Part I - Experiment and the Results

Part II - Verification of the Results.

Operational Definitions:

Results of our Learning Styles test consist of points placed in one of four quadrants. Each point represents one test result (Fig.1). Group perception of an individual's learning style was evaluated by defining a mean learning style (MLS) and a learning style vector (LSV) pointing to this mean. This was done as follows:

a) The MLS was calculated as the centre of gravity of points for each quadrant.
b) The LSV was constructed by joining the MLS to the origin.

   The magnitude of the LSV is equal to its length. The direction of the LSV is equal to the angle between the LSV and the x-axis (i.e. the AE-RO line) measured in the counter clockwise direction.

The magnitude of the LSV determines the extent to which the person conforms to a particular style of learning as determined by a quadrant. For example, a shorter magnitude indicates that the degree to which he conforms is less as opposed to a person whose magnitude is longer. On the other hand, the direction of the LSV determines the "inclination" of the individual with respect to the perception (CE-AC) or processing (AE-RO) lines. Figures 1 and 2 below show the LSV for two hypothetical individuals, both of whom are in the Imaginative quadrant of the LSI diagram.

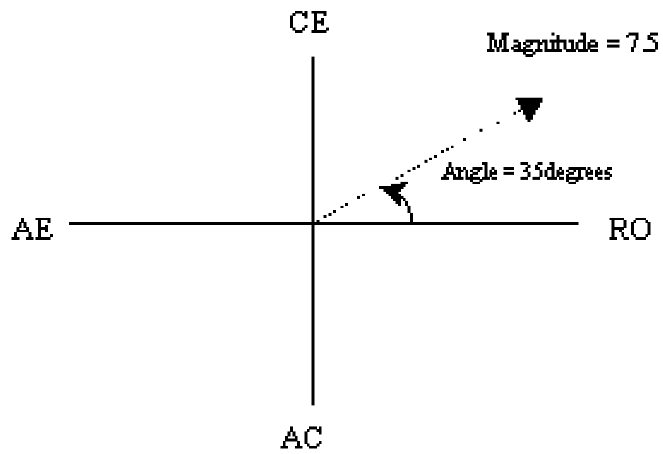


Figure 1. For person A

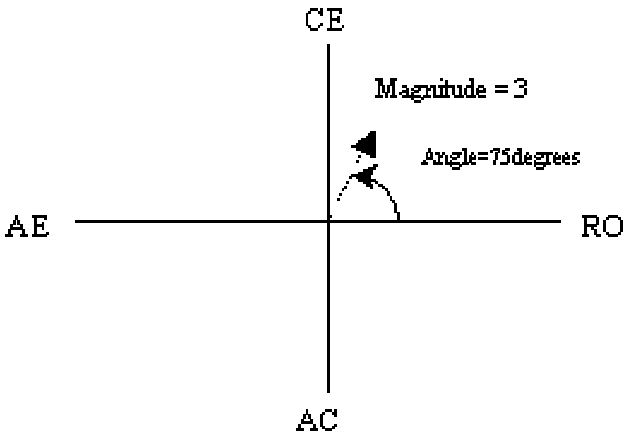


Figure 2. For person B

In Figure 1 the magnitude of the LSV is 7.5 whereas in Figure 2 it is 3.0. In other words, person A is more conforming in his learning style as compared to person B. Similarly, the angle in Figure 1 is 35 degrees whereas in Figure 2 the angle is 75 degrees. This means that person A is more inclined towards “Reflective Observation” (RO) while person B is more inclined towards “Concrete Experience” (CE).

Part I: Experiments and Results

Description of the Experiment

Six subjects code-named X, Y, Z, A, B and C were identified and were asked to respond to the LSI questionnaire. The same questionnaire was then given to six groups of people. Each group knew one or more of the subjects well. Each group was asked to respond to the questionnaire not as themselves but as one of the subjects they knew well. The group size evaluating each subject varied from 9 to 58 people.

Results

1) Subject X

150 questionnaires were distributed to various people who had interacted with person X in various capacities. They were asked to respond to the questionnaire as person X. The same questionnaire was also given to X for self-evaluation. 58 people responded to the questionnaire including X. The results are as follows:

Mean learning style of the total population on AC/CE dimension is -1.06 and AE/RO dimension is 3.0. X falls in the Dynamic quadrant. In other words, others perceive X to be a “Dynamic learner”. However, X perceives himself to be an “Imaginative learner”.

Table1. Mean & Standard deviation for the four quadrants

Mean & Standard deviation	Dynamic (no. of people=20)	Imaginative (no. of people=16)	Analytical (no. of people=13)	Precision (no. of people=9)
AC-CE continuum	-7.5	-6.12	7.5	9.8
SD	6.5	6.25	4.40	3.7
AE-RO continuum	10.5	-4.62	-3.5	9.1
SD	5.8	4.99	5.36	6.2

From Table 1 it is clear that the largest number of people have placed X in the Dynamic quadrant, followed by the Imaginative quadrant, then the Analytical and lastly the Precision quadrant.

Learning Style Vector: Figure 3 gives a graphical representation of the vectors in each quadrant. People who find X to be a Dynamic learner feel that X is inclined towards "Active Experimentation". The angle of the line is 35 degrees (to the AE side of the axis) with a magnitude of 12.35. On the other hand, people who find X to be an Imaginative learner, feel that he is more inclined towards "Reflective Observation". The angle of the line is 37 degrees (to the RO side of the axis) with a magnitude of 7.6. People who find X to be a Precision learner feel that he is inclined towards "Active Experimentation". The angle of line is 40 degrees (to the AE side of the axis) with a magnitude of 12.05. Lastly, people who perceive X to be an Analytical learner feel that he is more inclined towards "Reflective Observation". The angle of line is 25 degrees (to the RO side of the axis) with a magnitude of 8.31. The predominant style of conformance of X is Dynamic.

As the next step, we defined two subgroups of the original group that evaluated X. The first subgroup consisted of close associates of X, who had worked with him for 10 years or more. The second subgroup consisted of people who reported to X. This was done primarily to see whether there would be any difference in how the two subgroups perceive the learning style of X. The sample size selected in this case was purposefully smaller in number.

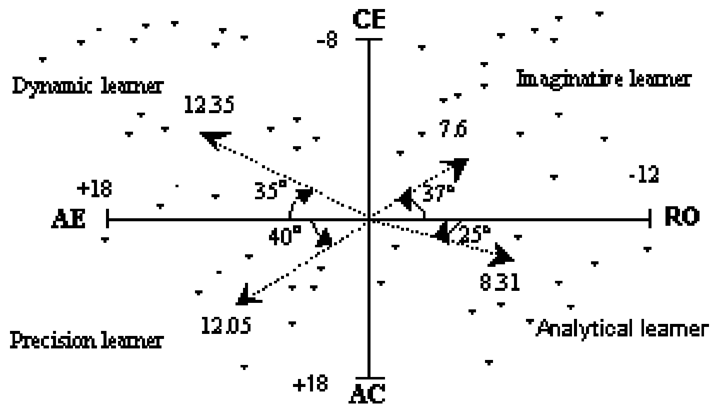


Figure 3. Subject X

<u>*Close associates of X:*</u> 9 people were identified in this subgroup. **Mean learning style** of the total population on AC/CE dimension is -1.33 and on the AE/RO dimension is 8.66. In other words, this group perceives X to be a Dynamic Learner.

<u>*People reporting to X:*</u> 10 people were identified in this subgroup. **Mean learning style** of the total population on AC/CE dimension is 0 and on the AE/RO dimension is 1.5. In other words, this group perceives X to be an Imaginative Learner.

Thus, people working under X perceive him as an Imaginative learner. Their perception matches with his own perception, i.e., with X's perception of himself. Whereas, people working close to X perceive him as a Dynamic learner, which does not match with his own perception. This is somewhat surprising as it may have been expected that people close to X should have perceived X the same way as X perceives himself.

2) <u>Subject Y:</u>

Mean learning style of the total population on AC/CE dimension is 6.00 and AE/RO dimension is -6.36. Average population perceives Y to be an Analytical learner. Y also perceives himself as an Analytical learner. If, however, we ignore Y's responses, then the shift moves from the analytical quadrant to the Imaginative quadrant. In other words, Y perceives himself as an Analytical learner while the others perceive him as an Imaginative learner.

Table 2 shows that equal number of people have placed Y in the Imaginative quadrant as in the Analytical quadrant.

Table2. Mean & Standard deviation for the two quadrants

Mean & SD	Imaginative (no. of people = 5)	Analytical (no. of people = 5)
AC-CE continuum	-2	12.8
SD	2.82	4.60
AE-RO continuum	-8	-10.4
SD	7.11	7.66

Learning Style Vector: Figure 4 gives a graphical representation of the vector. People who perceive Y as an Imaginative learner feel that he is more of a "Reflective observer". The angle of the line is 35 degrees (to the RO side of the axis) with a magnitude of 8.2. Similarly, people who find Y to be an Analytical learner feel that he is also inclined towards "Reflective observation". The angle of the line is 40 degrees (to the RO side of the axis) with a magnitude of 16.5.

The two groups agree that Y processes information by "Reflective Observation". However, the largest MLS is in the Analytical quadrant and matches Y's own perception of self.

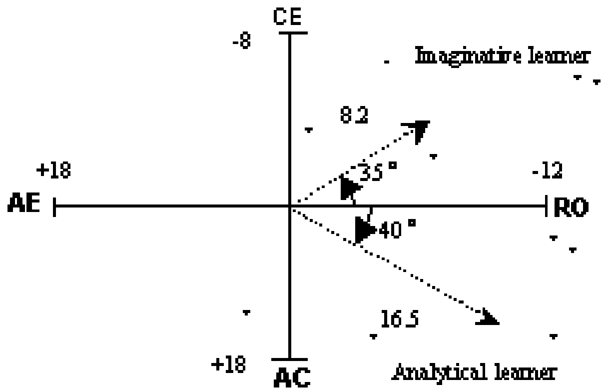


Figure 4. Subject Y

3) Subject Z:

Mean learning style of the total population on AC/CE dimension is 11.00 and AE/RO dimension is 7.4. Average population perceives Z to be a Precision learner. Whereas, Z perceives himself to be an Analytical learner.

From Table 3 it is clear that there are two distinct groups. One group who perceives Z to be a Precision learner and the other group perceives Z to be an Analytical learner.

Table 3. Mean & Standard deviation for the two quadrants

Mean & SD	Precision (no. of people = 7)	Analytical (no. of people = 3)
AC-CE continuum	14.33	7.33
SD	6.86	5.77
AE-RO continuum	11.0	0
SD	2.44	0

Learning Style Vector: Figure 5 gives a graphical representation of the vector. People who find Z to a Precision learner feel that he is more oriented towards "Abstract Conceptualisation". The angle of line is 52 degrees (to the AE side of the axis) with a magnitude of 18. People who find Z to be an Analytical learner also feel that he is inclined towards "Abstract Conceptualization". The angle of line is 70 degrees (to the RO side of the axis) with a magnitude of 7.33.

Both the groups agree that Z perceives information through "Abstract Conceptualization". However, the higher MLS is in the precision quadrant and is different from Z's self-perception.

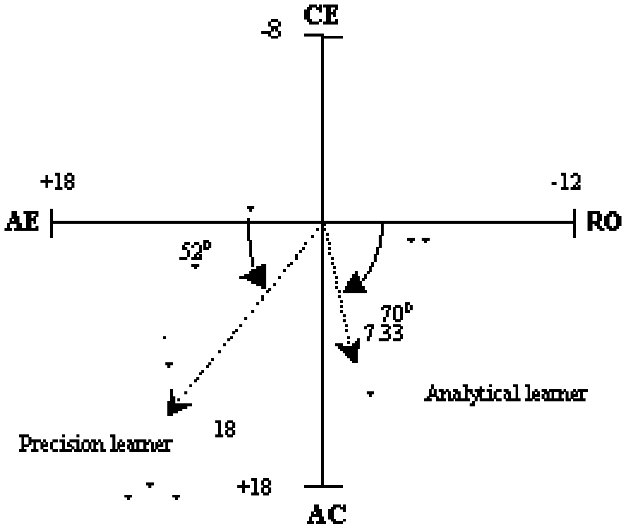


Figure 5. Subject Z

4) <u>SUBJECT A:</u>

Mean learning style of the total population on AC/CE dimension is -7.00 and AE/RO dimension is 4.54. Average population perceives A to be a Dynamic learner. However, A perceives herself to be an Imaginative learner.

Table 4 suggests that there are two distinct groups. One group perceives A as a Dynamic learner while the other group perceives A as an Imaginative learner.

Table 4. Mean & Standard deviation for the two quadrants

Mean & SD	**Dynamic** (no. of people = 5)	**Imaginative** (no. of people = 3)
AC-CE continuum	-17.2	-10
SD	6.2	8.48
AE-RO continuum	16.8	-7
SD	8.78	7.07

Learning Style Vector: Figure 6 gives a graphical representation of the vector. The MLS falls in the Dynamic quadrant and is different from A's perception of self. Both the groups agree on a high value for Concrete Experience but disagree on the AE-RO axis.

The predominant style of conformance is Dynamic.

5) <u>SUBJECT B:</u>

Mean learning style of the total population on AC/CE dimension is -2.36 and AE/RO dimension is -11.90. Average population perceives B to be an Imaginative learner. Though, B perceives herself to be an Analytical learner.

It is clear from Table 5 that there are two distinct groups. One group perceives B to be an Analytical learner and the other group perceives B to be an Imaginative learner.

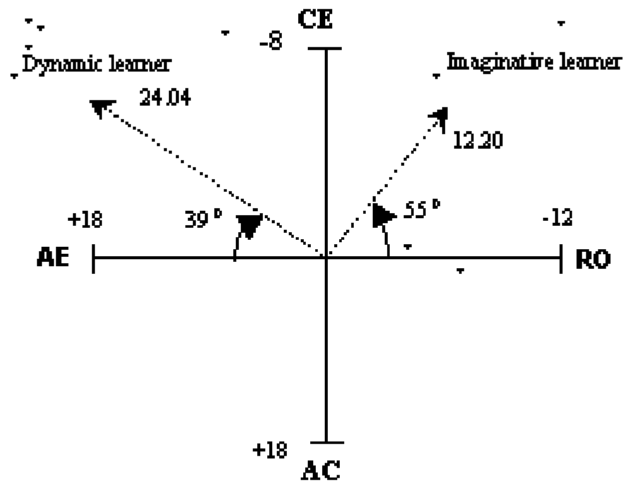


Figure 6. Subject A

Learning Style Vector: Figure 7 gives a graphical representation of the vector. Both the groups feel that B processes information through "Reflective observation". They disagree on the CE-AC axis. The MLS falls in the Imaginative quadrant which is different from B's perception of self.

Table 5. Mean & Standard deviation for the two quadrants

Mean & SD	Imaginative (no. of people = 9)	Analytical (no. of people = 2)
AC-CE continuum	-4	5
SD	6.48	1.41
AE-RO continuum	-9.66	-22
SD	5.83	2.82

6) <u>SUBJECT C:</u>

Mean learning style of the total population on AC/CE dimension is 4.61 and AE/RO dimension is 2.76. Average population perceives C to be an Analytical learner, though, C perceives himself to be an Imaginative learner. Table 6 suggests that there are two distinct groups. One group perceives C to be an Analytical learner and the other group perceives C to be an Imaginative learner.

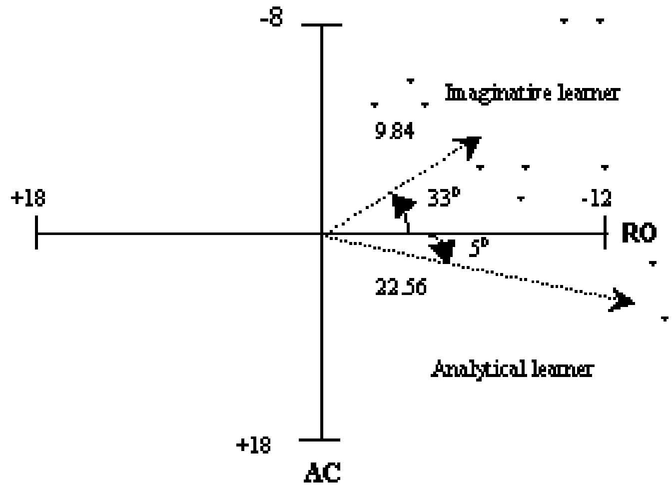


Figure 7. Subject B

Table 6. Mean & Standard deviation for the two quadrants

Mean & SD	Analytical (no. of people = 7)	Imaginative (no. of people = 4)
AC-CE continuum	12.66	-3.5
SD	*5.88*	*4.7*
AE-RO continuum	*-2.5*	*-2.5*
SD	*4.27*	*4.43*

Learning Style Vector: Figure 8 gives a graphical representation of the vector. Both groups agree on the AE-RO axis but disagree on the CE-AC axis. The MLS falls in the Imaginative quadrant.

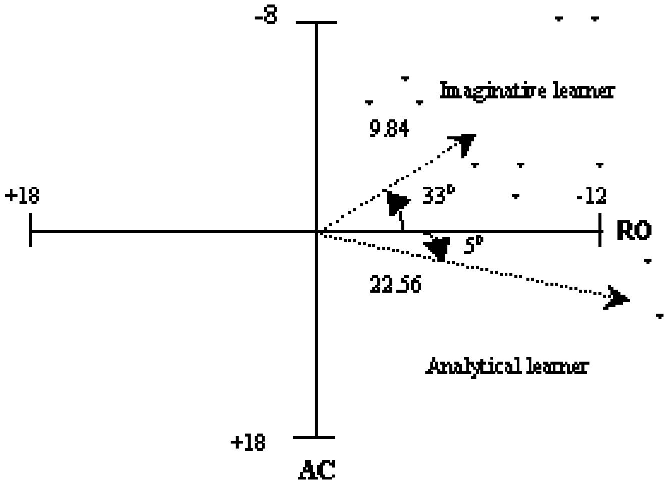


Figure 8. Subject C

Part II: Verification of the Results

Description of the Experiment

After conducting the first experiment as described above, we decided to conduct another experiment to check the validity of the test itself. Here, we presented each person who had marked the test for another person, with a description of the results. That is, if a person, P, had evaluated another person, A, as dynamic, we would present the standard description of a dynamic learner to P and ask him/her to compare this description against their intuitive perception of A according to the four options shown below:

Option 1: The description matches perfectly.

Option 2: The description matches to a great extent.

Option 3: The description matches somewhat.

Option 4: The description does not match at all.

This feedback form was given to the people who had responded for others and/or themselves. Each person compared the results for four other persons.

The same exercise was repeated on another group. However, of the four test results and corresponding descriptions that were given, two descriptions were purposefully selected to be incorrect. That is, if a person had marked a test such that the result evaluated to, say, analytical, he was given the description of the, say, dynamic style and asked to match that with his perception of the person he had marked for.

Results

In the first case, everybody marked either option one or option two. That is, they agreed perfectly or to a great extent with description of the person for whom they had responded.

In the second case, all respondents selected options one or two for the correct descriptions and options three or four for the incorrect ones.

We find these results highly supportive of Kolb's original descriptions (Kolb 1984b) of each learning style. All respondents were effortlessly able to identify the incorrect descriptions.

In fact, one is tempted to state that the results seem to indicate that while people may have misconceptions about what their own or other's learning styles may be, they are consistently clear about what the style is not!

Conclusions

The results give indicative answers to the questions posed earlier. For the cases studied:

1) There is, often, a difference between how we perceive our learning style and how others perceive our learning style.

2) Perceptions of self by self, and of self by others, fall in adjacent quadrants and never in diagonally opposite quadrants.

3) Responses are not scattered uniformly over the graph. On the contrary, there is a definite clustering of responses. These clusters are formed in the quadrant that is indicated through self-perception and in an adjacent quadrant.

4) The 360-degree technique used with a LSI test is found to be useful in determining how different groups perceive a person. For example, for evaluating the perceptions of subordinates, friends or family.

5) The LSI descriptions correctly reflect our perceptions of our own, or others, learning styles.

6) The "real" learning style is neither based on self perception or other's perception of self but is somewhere in between.

Speculative Comments

In every case that we studied, self-perception was different from the perception of self by others. What could this indicate?

a) Do we identify our learning style as it really is or as we would like it to be?

b) Do we have a perceived style, that is "what we really are" but we project another style that we want others to perceive us as? Is this process deliberate on our part? Do we have a private image that is core to us and a public image that we purposefully project in front of others? Maybe, the public self becomes operational only when we interact with others. The reason for this projection could be varied, for example, self-preservation, vulnerability, adaptation to the environment, etc.

Or

c) Is there a manifestation of two distinct personalities? Is this universal? If this is true, can a 360-degree technique measure the extent of the difference between these two personalities?

References

Cronbach, L & Snow, R. (1977). Aptitudes and Instructional Methods: A Handbook for Research on Interactions. New York: Irvington.

Dangwal, R., & Mitra, S. (1997). Construction and validation of a Learning Styles Inventory test for use in India. International Conference on Cognitive Systems (Vol 2), New Delhi, Allied Publishers.

Dun, R., & Dun, K. (1978). Teaching students through their individual learning styles: A practical approach. Reston, VA: Reston Publishing.

Keefe, J.W. (1979). Learning Style: An overview. In NASSP's Student learning styles: Diagnosing and prescribing programs (pp. 1-17). Reston, VA: National Association of Secondary School.

Keefe, J.W. (1989). Learning style profile handbook: Accommodating perceptual, study and instructional preferences (Vol. II). Reston, VA: National Association of Secondary School Principals.

Kolb, D. (1984a). Experiential learning: Experience as the source of learning and development. Englewood Cliffs, NJ: Prentice-Hall.

Kolb, D. (1984b). Learning Styles Inventory. Boston: McBer & Co.

Montgomery, S. (1996). Addressing Diverse Learning Styles through the Use of Multimedia, University of Michigan.

Mumford, A. & Honey, P. (1996). Using your learning styles, Honey, Maidenhead.

Myers, I. (1978). Myers-Briggs Type Indicator. Palo Alto, CA: Consulting Psychologists Press.

Reichmann, S.W., & Grasha, A.F., (1974) A rational approach to developing and assessing the construct validity of a student learning style scale instrument. Journal of Psychology, 87, 213-223.

Schmeck, R. R. (1983). Learning styles of college students. In R. Dillon & R. Schmeck (Eds.), Individual differences in cognition (pp.233-279). New York: Academic Press.

Witkin, H.A. (1954). Personality through perception: An experimental and clinical study. Westport, CT: Greenwood Press.