

# 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](mailto:dtmt@senet.com.au)

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

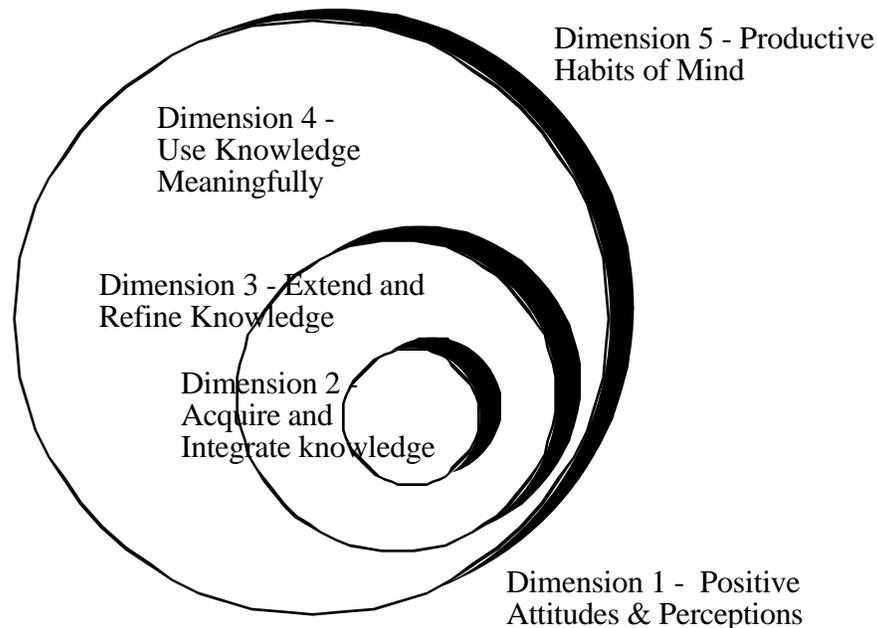
*In order to test the effectiveness of the implementation of the Dimensions of Learning Program in an Australian independent boys school, data from the Australian Schools Science Competition over a number of years from this school were analysed. Subjects were secondary school students from the school 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 :

$$Y_{ij} = \beta_{j0} + \beta_{j1} (\text{grade level}_{ij}) + \beta_{j2} (\text{exposure to DOL}_{j2}) + r_{ij}$$

In this equation  $Y_{ij}$  represents the performance of student  $j$  at grade level  $i$  and  $\beta_{j0}$  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  $\beta_{j1}$  represents the effect of the grade level and its parameter value in the bracket represents the particular grade level. The coefficient  $\beta_{j2}$  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  $\beta_{j1}$ , 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.

$$\beta_{j1} = \gamma_{01} + \gamma_{11} (\text{cohort1993}) + \gamma_{21} (\text{DOLI}) + \gamma_{31} (\text{IQ}) + u_{j1}$$

where  $\gamma_{01}$  represents the average level across the students of the coefficient  $\beta_{j1}$ ,  $\gamma_{11}$  represents the influence of being in the 1993 cohort of students,  $\gamma_{21}$  represents the influence of the Dimensions of Learning Program on growth, and  $\gamma_{31}$  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 an 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   BCBDAAABDCADDBCBABADDBCCBCDBDADDABCBCABDBDB

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 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
<b>Cohort group 1993</b>	Grade 9 no DOL	Grade 10 no DOL	Grade 11 DOL	Grade 12 DOL	-	51
<b>Cohort group 1994</b>	Grade 8 no DOL	Grade 9 no DOL	Grade 10 DOL	Grade 11 DOL	Grade 12 DOL	103
<b>Cohort group 1995</b>	-	Grade 8 DOL	Grade 9 DOL	Grade 10 DOL	Grade 11 DOL	98
<b>Cohort group 1996</b>	-	-	Grade 8 DOL	Grade 9 DOL	Grade 10 DOL	103
<b>Cohort group 1997</b>	-	-	-	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

$$Y = B_0 + B_1 * (\text{LEVEL}) + B_2 * (\text{DOL}) + R$$

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

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

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

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

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, the intraclass correlation can be calculated.

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

**Table 2. The final estimation of the fixed effects**

Fixed Effect	Coefficient	Standard Error	T-ratio	P-value
<b>For INTRCPT1,B0</b>				
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
<b>For GRADE LEVEL slope, B1</b>				
INTRCPT2,G10	0.263	0.031	8.6	0.000
DOL1, G11	-0.018	0.037	-0.5	0.624
<b>For DOL slope, B2</b>				
INTRCPT2,G20	0.106	0.059	1.8	0.036 <sup>a</sup>
IQ, G21	0.0055	0.003	2.0	0.043

<sup>a</sup> 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