Is student achievement really immutable?

A study of cognitive development and student achievement in an Oregon school district

Thesis submitted

by

Lorna Christine ENDLER BSc(Hons), PGCE, MEd(Hons)

in September 2004

for the degree of Doctor of Philosophy

in the School of Education

James Cook University

STATEMENT OF ACCESS

I, the undersigned, author of this work, understand that James Cook University will make this thesis available for use within the University Library and, via the Australian Digital Theses network, for use elsewhere.

I understand that, as an unpublished work, a thesis has significant protection under the Copyright Act and:

I do not wish to place any further restriction on access to this work.

Signature ________________ Date ________________

STATEMENT OF SOURCES - ELECTRONIC COPY

I, the undersigned, the author of this work, declare that the electronic copy of this thesis provided to the James Cook University Library is an accurate copy of the print thesis submitted, within limits of the technology available.

Signature

Date

STATEMENT OF SOURCES

DECLARATION

I declare that this thesis is my own work and has not been submitted in any form for another degree or diploma at any university or other institution of tertiary education. Information derived from the published or unpublished work of others has been acknowledged in the text and a list of references given.

Signature

Date

STATEMENT ON ETHICS

DECLARATION

The data for this thesis were collected as part of the regular operations of the Molalla River School District, Oregon, USA. Ethical considerations concerning the project therefore rest properly in the hands of the administrators of the Molalla River School District. The data were subsequently provided to me by the Director of Instruction of the Molalla River School District, and further analyses have been stored as anonymous records.

Signature: Date

STATEMENT ON THE CONTRIBUTION OF OTHERS INCLUDING FINANCIAL AND EDITORIAL HELP

IRA grants from James Cook University provided funding for travel in 2000 and 2001 to Portland, Oregon, from my home in California for seminar presentations and data collection. I was also awarded a JCU Completion Scholarship in September 2003. There was no other financial support for the study. Further contributions by others to this work are acknowledged overleaf.

ACKNOWLEDGMENTS

I am grateful to a number of people who have provided assistance with this work. Most importantly, I would like to thank my supervisor, Dr. Trevor Bond, for his wise and ever positive comments, and his unfailing encouragement. Thanks are also due to the Administrators of the Molalla River School District, in particular to Dr. Sandra Pellens the Director of Instruction, for allowing me to access the data analyzed in this study, and to Eric Bigler, late of Clackamas ESD, for sending me the data files over the period of the study. Joe Zenisek, of Molalla River High School, and Sallie Niggs, of Molalla River Middle School were invaluable school contacts. Lastly, I would like to thank my husband, John, for miscellaneous advice, and for his endless patience and good humour.

ABSTRACT

In the educational climate of the USA, where many question the possibility of effecting genuine change in national achievement outcomes, the Scientific Thinking Enhancement Project (STEP) was delivered to three cohorts of students from 1999 to 2002 in Molalla, Oregon. At the start, the mean age of Cohorts A, B, and C was 11+, 12+, and 13+ years. The purpose of my study was to investigate whether the STEP had enhanced these students' cognitive development and school achievement. The STEP incorporated strategies from a British intervention that had been shown to have a substantial effect on children's cognitive development and school achievement. Different test instruments were employed from those used in the British intervention, and the results of all tests were Rasch scaled. Cognitive change was estimated using *Bond's Logical Operations Test*, with pre-intervention performance profiles serving as cross-sectional controls. Statistical analyses revealed some enhancement of cognitive development compared with controls, with cognitive gains across the spectrum of starting level, irrespective of starting age and level of parent education. Statistically significant overall cognitive gains were found for Cohorts B (0.27 SDs) and C (0.55 SDs). Data from state-mandated tests in Mathematics revealed significant overall gains against controls for Cohorts A (0.51 SDs) and B (0.19 SDs). Cohort B students also made late-onset significant gains over peers who missed the STEP in 8th grade (BLOT 1.01 SDs and Mathematics 1.09 SDs). Cohort B females showed a significant overall gain in state Reading & Literature tests. There were no significant achievement gains against populations from non-project schools. A teacher survey showed general satisfaction with the STEP, but also revealed misconceptions about the intervention. Given that these teachers received little professional development, and did not deliver the entire intervention program, it is not surprising that the STEP did not yield results as strong as the original projects in the UK.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Is student achievement really immutable?

> .....schools bring little influence to bear on a child's achievement that is independent of his background and general social context.
>
> Colman & Associates (1966) p. 325

> The lines on a graph of average student performance are as flat as the surface of a frozen lake.
>
> Bracey (1987) p. 109

> Teaching-centered schooling has reached its upper limit: It doesn't get any better than this.
>
> Branson (1998)

Observers of national trends in student achievement in the United States over the past forty years have concluded that the level of student achievement is almost immutable, in spite of the school reform that has taken place over the same period (Peaker 1975; Bracey, 1987; Branson, 1998; Barton & Coley, 1998). However, international comparative studies of student achievement have provided a basis for questioning previous conclusions that school practices do not influence student achievement (Suter, 2000). Studies of curriculum coverage in mathematics and science carried out during the Trends in International Mathematics and Science Study (formerly known as the Third International Mathematics and Science Study, TIMSS) have identified large differences between countries in regard to the coverage of science and mathematics topics (Schmidt, Jorde, Cogan, Barrier, Gonzalo, Moser, Shimizu, Valverde, McKnight, Prawat, Wiley, Raizen, Britton, & Wolfe, 1996; Schmidt, McKnight, Cogan, Jakwerth, & Houang, 1999).

The various studies of student achievement that have been conducted since 1965 have influenced educational policy and the direction of research in the United States. Guided by early studies such as the 1966 Coleman Report, that stressed the importance of family background, policy makers tended to concentrate on the plight of minority groups and improving equality of opportunity in schools, rather than on other educational practices within schools. More recently, frameworks of national and state content standards have been formulated as a means of improving student achievement. However, these large-scale school reforms do not appear to have been successful in enhancing the achievement of the majority of students. Indicators of student achievement such as the National Assessment of Educational Progress have shown little change over the period from 1973-1996 (Campbell, Reese, & Dossey, 1996; Barton & Coley, 1998). The message that the public and politicians receive from these national data sources is that little can be done to change student achievement. The scarcity of research into other educational practices during the same period has compounded the view that student achievement is virtually immutable. However, the results of programs such as the Trends in International Mathematics and Science Study (TIMSS) and the Program for International Student Assessment (PISA) are now directing attention to how schools might influence student achievement.

The specific goal of TIMSS was to identify and assess the components of mathematics and science stressed in each participating country. Analysis of the data from the 1994-5 TIMSS suggested that the content and organization of mathematics and science curricula, and how these subjects are presented in the classroom, might influence student achievement (Mullis, Martin, Beaton, Gonzales, Kelly, & Smith, 1998). Results indicated that a lack of focus in the curricula and the extremely variable pattern of school offerings contribute to the relatively low U.S. achievement levels in these international comparative studies. The high diversity of ethnicity in the United States does not appear to have been an important factor influencing the nation's ranking in TIMSS. Scores for the white only students at the end of secondary school place the United States in the lowest one third of all the participating countries, with the performance of minority groups having little effect on the country's overall position in this study. High school physics students, the country's most highly selected and potentially top achievers, were in fact amongst the lowest performers of 17 participating countries (Mullis, Martin, Beaton, Gonzales, Kelly, & Smith, 1998).

A recent report released by the National Center for Educational Statistics (NCES on-line, 2003) focused on classroom practice as seen in videotapes of 8th grade mathematics classrooms in the U.S. and in six top-performing nations that participated in the TIMSS videotape studies: Hong Kong, Japan, the Netherlands, the Czech Republic, Australia and Switzerland. The U.S. was ranked the lowest of these countries in both the 1995 and 1999 TIMSS surveys. The videotape report offers many insights into how mathematics is taught in the different countries, but presents few conclusions. While the report showed some general features among the countries, there was considerable variation in the teaching of 8th grade mathematics. Distinctions included the introduction of new content, the coherence across mathematical problems and within their presentation, the topics covered and the procedural complexity of the problems, and classroom practices regarding individual student work and homework.

The results of the Organization for Economic Cooperation and Development's Program for International Student Assessment (PISA) show only average performance in reading, mathematics and science for 15 year old U.S. students, when compared with participants from 32 countries (NCES on-line, 2002; OECD on-line, 2003). PISA is a relatively new system of international assessment that focuses on literacy skills in reading, mathematics and science. The purpose of PISA is to represent the overall yield of learning that 15 year olds have acquired as they near the end of mandatory schooling.

The low ranking of the USA in international comparative studies such as TIMSS and PISA has stimulated many school districts to take a closer look at what is occurring in mathematics and science classrooms. It was in this general context that the Molalla River School District in Oregon, USA adopted an intervention titled the Scientific Thinking Enhancement Project (STEP) in 1999. The STEP project used strategies taken directly from the Cognitive Acceleration through Science Education (CASE) program, that has been shown to raise student achievement levels in other parts of the world.

Secondary science curriculum materials in the UK require students to use formal thinking (Shayer & Adey, 1981; Adey & Shayer, 1994). This appears to be also true in other countries such as Australia (Macdonald, 1994; Endler & Bond, 2001; Stanbridge, 2002), Pakistan (Iqbal & Shayer, 1995), and Denmark (Holbech & Thomsen, 2002). It follows that a mismatch frequently occurs between intellectual ability and the demand of the curriculum material for the many students who are still operating at the concrete level of thinking in secondary school. The Piagetian interpretation would be that students need to use formal operational thinking to succeed in assessment tasks that include complex scientific or mathematical concepts. Higher order thinking skills are essential for the manipulation of formulae, the design of scientific experiments, and making the necessary connections between concrete experimental data and abstract scientific theory. From this perspective, the study of science or mathematics becomes more than just a challenge to a student whose thinking has not yet reached the formal operational period of cognitive development. Achievement in science, and in other subjects, is likely to be related to cognitive development, and enhancement of cognitive development would therefore be likely to improve student achievement at the high school level.

The success of the CASE strategies in improving cognitive development has been demonstrated convincingly in England (Adey & Shayer, 1990; Shayer & Adey, 1992 a & b; Shayer, 1999; Adey & Shayer, 2002), in Pakistan (Iqbal & Shayer, 2000), in Australia (Endler, 1998 a & b; Endler & Bond, 2001), in Denmark (Holbech & Thomsen, personal communication 2002), and in Finland (Kuusela & Hautamäki, 2002). The British CASE research demonstrated a positive relationship between the data from results of the General Certificate of Secondary Education and cognitive tests (Adey & Shayer, 1994, 2002), suggesting an association between cognitive development and student achievement. The success of CASE is attributed to its strategies that actively promote the development of formal thinking in school students (Adey & Shayer, 1990, 1994).

In the United States there is considerable interest in programs such as CASE that claim to enhance student achievement, not least because schools are now held accountable to demonstrate growth if they are to attract national and state funding. I was invited by the Director of Instruction in the Molalla River School District to assist with a seminar presented by Dr. Trevor Bond to the STEP teachers in May 2000, and again in April 2001, because of my experience with a CASE project in Townsville, Australia. These visits to Oregon from my present home in California were funded by IRA grants from James Cook University. When I expressed an interest in analyzing the results of the STEP project, it was agreed that I would be given access to the results of *Bond's Logical Operations Test*

(Bond, 1976) that was used to estimate the students' cognitive development, as well as to the state achievement scores of the participating students.

My role in the STEP project has been that of an informed observer of the outcomes of the program, situated distantly from where it was taking place. I did not initiate the STEP project, nor did I have any control over the testing procedures that provide the data for my analysis. Both the intervention program and the testing occurred quite independently of my involvement. Indeed, it was after the project had been organized that I became involved, and both the intervention and the testing would have gone ahead whether or not I had expressed an interest in analyzing the data. Ethical considerations concerning the project and the analyses of data therefore rest properly in the hands of the administrators of the Molalla River School District.

Decisions about which STEP activities to use with which grades were made by Dr. Pellens, the Director of Instruction in the Molalla River School District and the STEP teachers. Decisions about which tests to give, and advice to teachers as to when testing would occur, came from Dr. Pellens and Dr. Trevor Bond. These decisions were made before I became directly involved with the STEP project. However, their decisions were strongly influenced by my MEd research on a CASE project in Australia (Endler & Bond, 2001).

Dr. Pellens first became interested in the idea of using CASE when she saw our results presented by Dr. Bond at a workshop in the USA. She was impressed by the success of the Australian CASE project and with the use of repeated BLOT tests to estimate cognitive development. According to Dr Bond, she wrote the grant proposal for

STEP with our report in front of her and with my later participation in mind. As it turned out, STEP did not entirely follow our Australian model. Dr. Pellens had little control over what the STEP teachers actually did, and her advice was often ignored.

Initially, the only data Drs. Pellens and Bond planned to collect were from the BLOT testing. I requested the school achievement data and had to constantly follow up when they did not send it to me (this was also the case with the BLOT data sets). It required a lot of perseverance and detective work to build up the master files. If I had used the data as sent, I would missed about 30% of the cases. I also sought comparison data from two other school districts.

I designed the data analyses used in the study. Clearly Dr. Bond influenced me in using the Rasch model, which I had also used in my Masters study. His main input for the PhD analysis was in the issue of anchoring. All the statistical tests and procedures for using the BLOT and achievement data, and showing the relationship between them, were my own conceptualization. Dr. Bond's perspective on my part in the design of the STEP project is that what I did not directly suggest was often taken from the model used in my Masters work.

It was never intended that the STEP would be a replication of the original CASE studies in the UK, nor, indeed, that it would be a research project of any kind. The administrators of the Molalla River School District adopted the CASE intervention with the intention of enhancing the scientific inquiry skills of their students. The goal was for students and teachers to benefit, and the idea of setting up concurrent experimental and control classes was not considered.

It fell to me to make the best use of the available data to detect any possible changes in cognitive development and student achievement over the 32- month period of the study. In the absence of control data from concurrent classes, I made comparisons between the results of the STEP students and a student sample from the Molalla River School District in 1999, before the intervention began. This seemed a particularly reasonable solution, given that many states in the USA allocate funding to schools based on the improved student performance in the state-mandated tests of the current year compared with the previous one. Indeed, improved student performance compared with the pre-intervention sample was the very outcome by which the administrators of the Molalla River School District would judge whether to continue with the STEP intervention.

The primary goal of my study was to discover whether the STEP would enhance the cognitive development of the Molalla River students who were exposed to the intervention. Previous research has suggested that periods of more rapid cognitive change can be detected in the early adolescent years (Shayer, Küchemann & Wylam, 1976; Shayer & Wylam, 1978; Shayer & Adey, 1981; Endler & Bond, 2001). A question to be answered in this study was whether periods of more rapid cognitive change might be detected in the STEP students over the 32-month period of data collection.

Much of the post-Piagetian research has shown that there is a great deal of variation in the process of cognitive development. The nature of the schooling that children receive appears to be an important factor affecting this aspect of adolescent development (Kuhn & Angelev, 1976; Rosenthal, 1979; Shayer & Adey, 1981). Furthermore, teachers and parents are aware of the existence of early and late developers. A further question for this study was whether children in the Molalla River School District would also show differing rates of cognitive change.

Adey (1992) reported differences in the pattern of cognitive development of young male and female adolescents, attributed to the later onset of adolescence in males. The proportion of girls reaching formal operational thinking did not increase after the age of 14 years, whereas this effect was seen later, at 15 years, in boys. It was therefore important to compare the patterns of cognitive development of male and female students in the STEP cohorts.

Adey and Shayer (1994) reported increased cognitive development from the full range of pre-intervention students. They found that the students who made the greatest gains were not restricted to those who had much ground to make up, nor to those who were more able, and therefore perhaps more ready to move on to formal thinking; a finding confirmed by Endler & Bond (2001). Whether the STEP project affected the full range of students was another interesting question to be investigated with the Molalla data.

The second major goal of this study was to search for evidence that any cognitive gains were accompanied by improved student achievement. Shayer (1999) reported evidence of the long term effect of CASE on students' school results in science, mathematics and English in the UK. This trend was supported by our Australian CASE study, in spite of using different indicators of student achievement (Endler & Bond, 2001, 2002). In the current project, data collected annually by the Clackamas Test Scoring & Reporting Services were used to follow the achievement of individual Molalla River students in the three cohorts, who began the STEP project in 6th, 7th and 8th grades in

September 1999, over the 32-month period to June 2002. Data were also made available to compare the progress of STEP students with students who did not experience the STEP intervention in two other Oregon school districts with similar socioeconomic backgrounds to that of Molalla River.

Adey and Shayer (1994) described differences in the student achievement of male and female CASE students. Further questions to be addressed in this study were whether a relationship existed between achievement and the sex of STEP students, and whether the age of commencement of the intervention was important. Shayer and Adey (1993) found that girls were favoured by an earlier exposure to the CASE strategies at age 11+. The three cohorts of Molalla River students who began the STEP program in 1999 were then aged 11+, 12+ and 13+ respectively, and their progress was compared with that of their age-equivalent peers in the United Kingdom.

A summary of the goals and specific research questions:

Goal 1: To investigate whether cognitive development was enhanced by the STEP intervention in the Molalla River School District.

Specific research questions:

1.1 Has the cognitive level of STEP students changed over the period of the study?

1.2 How is the pattern of cognitive change of STEP participants different from that of the Molalla River school sample prior to the intervention?

1.3 What is the relationship between cognitive change and level of parent education for STEP students ?

1.4 What is the relationship between cognitive change and cognitive developmental level at the start of the STEP intervention?

1.5 What is the relationship between cognitive change and the age at which the STEP intervention commenced?

1.6 What is the pattern of cognitive change for individual students?

1.7 How is the pattern of cognitive change different for males and females?

1.8 How is the pattern of cognitive change different for students who receive only part of the intervention?

Goal 2: To discover whether cognitive change is accompanied by change in the Mathematics and Reading & Literature achievement of STEP participants.

Specific research questions:

2.1 Has the achievement of STEP participants in Mathematics and Reading & Literature changed over the period of the study?

2.2 How are the STEP participants' patterns of achievement in Mathematics and Reading & Literature different from those of the Molalla River school sample prior to the intervention?

2.3 What is the relationship between change in achievement level in Mathematics/ Reading & Literature and level of parent education for STEP students?

2.4 What is the relationship between change in achievement level in Mathematics/ Reading & Literature and STEP participants starting achievement levels?

2.5 What is the relationship between change in achievement level in Mathematics/ Reading & Literature and age at which the STEP intervention commenced?

2.6 How is the pattern of student achievement in Mathematics and Reading & Literature different for males and females?

2.7 How is the pattern of student achievement in Mathematics and Reading & Literature different for children who receive only part of the intervention?

2.8 How is the pattern of student achievement of STEP participants in Mathematics and Reading & Literature different from that of peers in comparison schools?

2.9 What is the relationship between cognitive level and achievement in Mathematics, Reading & Literature and Science for STEP participants?

The composition of the student body of the Molalla River School District was over 90% white. Therefore, questions regarding the effect of ethnicity on student achievement and cognitive development unfortunately remain outside the scope of this study.

Chapter organization

The literature review that follows in Chapters 2 and 3 explores the theoretical issues and empirical evidence central to this study. Chapter 2 gives an overview of cognitive developmental theory, with particular reference to the contributions of Piaget, Vygotsky and Feuerstein. This is followed in Chapter 3 by a review of the literature that informed the original CASE project, and a summary of the results of the various CASE studies.

The methods used in this research are discussed in Chapter 4. Information is presented about the Molalla River School District, the student sample, the intervention and the testing procedure. Chapter 4 also reviews the analytical tools used in the research and provides a rationale for the application of Rasch modelling in the analysis of the cognitive data collected for this study.

Chapters 5 and 6 present the results of the analyses conducted to fulfill the goals of the study. Chapter 5 is a review of the results of the analyses of the data collected on cognitive development, and the results of the analyses of student achievement data are described in Chapter 6. The new skills acquired by the STEP students are considered in Chapter 7, and Chapter 8 presents the results of a STEP teacher questionnaire. A summary of the results, conclusions, and a discussion of the implications of the study follow in Chapter 9.

CHAPTER 2

Cognitive developmental research

Cognitive development is regarded as a controversial topic by some educational researchers, particularly those who favour the influence of 'nurture' over 'nature'. That humans develop physically cannot be challenged, so why should the notion of cognitive development be so problematic for some educationalists?

During the period from 1930-1970 Jean Piaget, Bärbel Inhelder and their colleagues in Geneva provided empirical evidence that human cognitive functioning develops in an age-related fashion. Neurological studies (Epstein, 1986; Hudspeth and Pribram, 1990) have confirmed that humans experience a series of critical growth periods, during which brain structure becomes increasingly well organised, culminating in the finalization of the prefrontal lobes, an important site for logical and control systems during adolescence. Research on human intelligence has revealed a similar story. Cattell (1963) revealed that measures of fluid intelligence (cognitive processing ability) grow steadily, but not uniformly, to a plateau at about 16 years of age. Studies of second language acquisition also suggest that there are critical periods in development that can be related to localized brain function (Johnson & Newport, 1993).

Teachers are also aware that children's understanding of concepts develops in a stage-like way, becoming progressively more sophisticated as they get older. Performance on tests, reports, and other forms of assessment are the indicators that inform teachers and parents about student progress. If the assessment items are based mainly on content, then memory and recall become important factors in achievement. However, if the test items require higher order process skills, then it will be only those children who have developed more sophisticated cognitive processes that will achieve high results.

In normal development and schooling, many children will respond to opportunities to develop their cognitive processes. For others, the opportunities might not arise, especially in educational systems that concentrate on recall of facts rather than on understanding. If such opportunities come at a time when the gap is too great between a child's cognitive level and the level of cognitive processing demanded by the curriculum, then the resulting cognitive development is unlikely to occur.

Cognitive development is not synonymous with learning, but Adey and Shayer (1994) recognised that a considerable overlap exists between the two. Tanner's learning-development spectrum (Figure 2.1) illustrates the idea that it is in the area of overlap that the connection between learning and development becomes apparent (Tanner, 1978). Movement from left to right in the figure indicates an increasing realisation that learning is dependent at least to some extent on development, a view that has in part given rise to teaching strategies more sophisticated than those further to the left.

The Learning end of the spectrum is marked by rote learning, by which multiplication tables or foreign vocabulary can be effectively memorised, using frequent repetition and mastery learning. In the behaviourism described by Skinner (1976), the learner makes no connection between what is learned and prior knowledge, and the emphasis is on the material to be learned rather than on the learner. Concept learning takes place when newly gained knowledge is connected to preestablished concepts, as in the learning hierarchies of Gagne (Gagne, 1965). The incorporation of new learning into a network of existing concepts is fundamental to Ausubel's meaningful learning (Ausubel, 1968). At this point in the learning-development spectrum, the emphasis moves from the material to be learned towards a consideration of the learner as an individual who has prior knowledge.

THIS IMAGE HAS BEEN REMOVED DUE TO
COPYRIGHT RESTRICTIONS

Figure 2.1 Tanner's learning-development spectrum (after Adey & Shayer, 1994)

The poles of the spectrum also serve to illustrate the distinction between instruction and intervention, with the latter meaning the manipulation of learning experiences to maximise cognitive development. Feuerstein, Rand, Hofman & Miller (1980) designed an intervention strategy known as Instrumental Enrichment (IE) to enhance the learning potential of socioculturally disadvantaged adolescents. Feuerstein's strategy was one of mediated learning, based on Vygotsky's (1978) idea that effective learning requires mediation by an adult or more capable other. In the case of IE, mediated learning involved the teacher acting as afacilitator of student learning. Vygotsky emphasised the importance of the social dimension in learning. He defined the zone of proximal development as the distance between the actual developmental level of the learner as determined by independent problem solving, and the level of potential development of the learner as determined through problem solving under adult guidance or in collaboration with more capable peers.

At the extreme Development end of the spectrum is located Rousseau (1762), who argued that a child's education should be left to nature, guided only by natural curiosity and with minimum intervention. Piaget's place close to that end is appropriate because part of his work has lead to the realisation that cognitive development can be enhanced if learners are challenged to reach their maximum potential.

Shayer (1999) attributed the theory base of the Cognitive Acceleration through Science Education (CASE) program to Piaget, Vygotsky and Feuerstein. Shayer first met, as he calls it, the '12 year gap' (between the best and worst of entrants to secondary school) when teaching Chemistry in London in the 1960s. This inspired him, first to understand it, and then to change it (Shayer & Adey, 2002). Shayer realized that the cognitive demand of the curriculum was beyond the cognitive developmental level of many of his students. Changing the cognitive demand of the Chemistry syllabus for the General Certificate of Education was not an option. However, Shayer recognized that changing the cognitive developmental level of children might be possible, if they were provided with opportunities to acquire the skills necessary for the development of new cognitive processes. As a result of this work, Adey and Shayer (1994) demonstrated that a close relationship exists between cognitive development and student achievement. It was to the work of Piaget that they first went to develop lessons based on children's levels of cognitive development.

Piaget's genetic epistemology

In the early part of the twentieth century, the prevailing view of the child's cognitive activity was that it was the same as that of adults, merely becoming more efficient with use. Piaget challenged this perspective by his claim that children's thinking passes through a series of developmental stages that progressively show greater sophistication. His idea that a child thinks and learns in qualitatively different ways during particular developmental periods was revolutionary at that time. In their book *The Growth of Logical Thinking from Childhood to Adolescence* (GLT), Inhelder and Piaget (1958) outlined the transition of children's thinking from the concrete operational period of childhood to the formal operational period of adolescence. The thinking of a child in the concrete operational stage is characterized by inductive reasoning limited by personal experience in the concrete world, whereas the formal operational adolescent can think effectively in the abstract mode. Formal thought can dissociate general ideas or concepts from the contexts in which they were learned and, therefore, does not require specific concrete cues in order to trigger the recall and use of these general principles. Concepts are intellectually manipulated by integrating them into universal generalisations or by taking these generalisations back to their first principles. Furthermore, formal operational thinking is hypothetico-deductive; the student is able to conceive of new ideas, concepts, hypotheses or principles, explore their implications and then test for their validity.

Piaget and Inhelder developed the revised clinical method (*méthode clinique*) to discover how the thought processes of the adolescent differ from those of a child. Piaget's original clinical method relied heavily on language, the child giving answers to questions posed by the examiner. The new method allowed the child to express ideas by manipulating objects, as well as by verbal communication. Typically, a number of children or adolescents were presented with apparatus such as a pendulum or a balance, and were required to explain how it operates. Pairs of investigators conducted each interview: the *secretaire* keeping a detailed record of the subject's activities, while the *interrogatrice* asked clarifying questions, in order to find out the extent to which the child could construct classes, relations, and so on, in a variety of contexts. The subject was effectively being asked to behave as a scientist by designing a series of experiments, observing the results and drawing logical conclusions. Inhelder is listed as one of the pair of investigators in many of the protocols now filed in the Archives Jean Piaget in Geneva, however none refer to 'le patron Piaget' as a participant (Bond, 1994). Bond compared published protocols with the typed originals and Piaget's hand written manuscript for GLT. This comparison revealed that considerable evidence was lost in transcription, with the result that the published accounts often give a misleading view of the *méthode clinique*. Bond concluded that there are insufficient details of procedures and questions in the published accounts to allow an exact replication of many Piagetian tasks.

Using the revised clinical method, Inhelder effectively presented adolescents with investigations that were related to the complex thinking that underlies the mastery of many scientific concepts. The various chapters of GLT illustrate the adolescent development of the major reasoning patterns required for effective thinking in the biological and physical sciences.

Inhelder (1954) had previously observed that adolescent mental operations differ from those of younger children. She deduced that adolescent thought reaches a higher degree of equilibrium, becoming more flexible and effective, so that adolescents are able to deal with complex reasoning problems. Adolescents develop the ability to imagine the many possibilities inherent in a situation, comprehending hypothetical propositions and compensating mentally for transformations in reality.

Piaget's earlier theoretical work (1953) resulted in two logical models, the 16 binary operations and the INRC transformations, that are used to define the structure of formal operations. The 16 binary operations model situations involving two factors, for example pendulum length and weight, where each factor has two values (*e.g.* long/short or heavy/light). An illustration of the binary operation known as complete affirmation would be the observation that a pendulum of heavy or light mass can swing with either a low or high frequency, with the conclusion that weight makes no difference to the frequency of oscillation. If the child being interviewed also concluded from other observations that length determines the frequency of oscillation, then another of the 16 binary operations, reciprocal implication, would have been demonstrated. The attainment of each of the 16 binary operations is determined by whether the adolescent is able to derive the proper logical relations from among the factors involved in the experiment. These logical relations are usually called functions.

Following his analysis of functions, Piaget (1953) formulated a second model of adolescent development, known as the INRC four-group, that describes how adolescent reasoning is used to manipulate the conclusions from any experiment. The INRC four-group specifies four cognitive transformations that adolescents can use to manipulate functions, namely identity (I), negation (N), reciprocity (R) and correlativity (C). N is the inverse operation of the four-group; N and R are both forms of reversibility or ways of reversing the operations of thought. I is an identity operator, which, when applied to a function I, leaves the function unchanged. C changes the conjunction operation to disjunction, and vice versa, but leaves everything else unchanged (Ginsburg & Opper, 1988).

The equilibrium of a simple beam balance provides a useful illustration of the INRC 4-group (Baldwin, 1967). If the equilibrium of a balance with equal weights at equal distances is disturbed by adding a further weight to one side, the equilibrium can be regained in several ways. Firstly, the added weight could be removed in the same way in that it was added to restore the starting position. Another possibility would be to add an identical weight to the lighter side. Thirdly, the heavier weights could be moved along the beam closer to the pivot than the weights on the lighter side. The last two strategies would restore the equilibrium, but not the starting arrangement.

An adolescent who has attained formal operational thinking would correctly conclude from experimenting with a balance that both weight and distance from the fulcrum are variables that affect equilibrium, and so determine whether balance is achieved (Inhelder and Piaget, 1958). The nature of the inverse relationship between the two variables (distance and weight) would be revealed if the adolescent discovered that a small weight combined with a great distance is equivalent to a large weight combined with a small distance. If the child simultaneously increased both the weight and the distance on one of the arms of the balance during the experiment, and discovered that this had no effect on equilibrium, then the child would have shown understanding of I (the identity operator that leaves a function unchanged). Reducing the distance while increasing the weight, diminishing the weight while increasing the distance, or diminishing both, would demonstrate N (the inverse operation). In discovering that a change in weight can be compensated by a change in distance, the child would have shown understanding of reciprocity (R).

Piaget argued that the period of formal operations could be subdivided into an early sub-stage, commencing at about 11 years, and a later stage, commencing at about 14 years. Inhelder and Piaget (1958, p. 347) said:

> [A]fter a phase of development (11-12 to 13-14 years) the preadolescent comes to handle certain formal operations (implication, exclusion, *etc*.) successfully, but he is not able to set up an exhaustive method of proof. But the 14-15 year old adolescent does succeed in setting up proofs (moreover, spontaneously, for it is in this area that academic verbalism is least evident). He systematically uses methods of control which require the combinatorial system-*i.e.*, he varies a single factor at a time and excludes the others ("all other things being equal"), *etc*.

Piaget is not very explicit about how the final stage of formal operations is attained. He suggested that neurological change occurs around the age of puberty and that this provides the physical basis for the transition from concrete to formal operations. A facilitative social environment with opportunity for the child to experiment with objects, and resulting internal cognitive reorganization are also seen by Piaget as crucial to the development of formal structure.

Inhelder's close involvement in building the theory of formal operations was acknowledged by the authors of GLT in the preface to the book (p. xxiii):

> In other words while one of us was engaged in an empirical study of the transition in thinking from childhood to adolescence, the other worked out the analytical tools needed to interpret the results. It was only after we had compared notes and were making final interpretations that we saw the striking convergence between the empirical and analytic results. This prompted us to collaborate again, but on a new basis. The result is this present work.

Bond (2001) traced and identified the individual contributions that Inhelder and

Piaget brought to the theory of formal operations. He argued that it was the result of Inhelder's research into Induction that was instrumental in changing Piaget's prior conceptions of mature thinking. Bond's translation of Inhelder's 1954 *Bulletin de Psychologie* paper reveals her profound understanding of the features of formal thinking, particularly in classroom science contexts. Bond concluded (p. 209):

> While I am sure this paper does not present a theory of formal operational thinking..... it does provide a systematic description of adolescent problem-solving behaviour that will appeal to secondary school teachers and others who work with adolescents".

Prior to their collaboration on GLT, the comprehensive descriptions and explanations of formal adolescent thinking were found first to be in Inhelder's published work, rather than in Piaget's. It would appear that Inhelder's contribution to the theory of formal operations might have been underestimated by those who assumed that it was Piaget who played the dominant part in building the theory of formal operations.

Vygotsky's social construction of reasoning

Increased interest in the sociocultural aspects of cognitive development has prompted researchers to explore how language, social interactions, and contexts contribute to cognitive development. The direction of this inquiry has been guided by Vygotsky's theory that social relationships underlie all higher mental functions. It follows that a child will internalize an experience only after transformational processes, that take place first between people, interpsychologically, before being directed inward, intrapsychologically (Vygotsky, 1981). An example of this transformation occurs when dialogue with others becomes internalized to become part of an individual's inner thoughts (Wertsch, 1985). Vygotsky also claimed that one can only understand the higher mental functioning of an individual by examining the preceding sociocultural context (Vygotsky, 1978).

Vygotsky viewed teaching and learning as social processes, and his zone of proximal development illustrates how teaching and learning interact. The zone of proximal development (ZPD) is defined as the distance between what an individual can accomplish alone and what is possible when assistance is given by a more capable collaborator (Vygotsky, 1978). The method used to determine this social component of learning begins with the psychologist administering a standard test, such as the Binet intelligence test, from which the mental age of the subject is estimated. The child is then taken through some of the easier incorrect items by the psychologist, who discusses the problems and gives various hints on their completion. The child is now able to solve more of the test items with the assistance of the psychologist , and so a new mental age can be calculated. The difference between the two test scores (with and without the mediation of the psychologist) represents the ZPD of the child (Shayer, 2002).

In the classroom setting, a more advanced peer might assist the development of a child by prompting, modelling, explaining, asking leading questions, discussing ideas, providing encouragement, or keeping the attention on the learning context (Carter & Jones, 1994; Jones & Carter, 1994). Contexts for such assistance might include peer tutoring, cooperative learning, and sibling relationships (Forman & Cazen, 1985). Furthermore, Vygotsky claimed that within the zone of proximal development, language and other social interactions between individuals enhance learning. Vygotsky's theories of socially mediated learning were extended by Rogoff (1995) to include cognitive apprenticeship and participatory approbation. Rogoff described cognitive apprenticeship in terms of an expert-novice relationship between the participants, whereas she considered cognitive development to occur mutually for the interacting pair in participatory approbation .

Signs and tools as mediators of knowledge also play an important part in Vygotskian theory (Vygotsky, 1978; Kozulin, 1990, 1998). Signs can include internalised ideas, beliefs and concepts; as well as symbols such as the alphabet, language and numbers, that are both externally and internally oriented. Tools, such as the memory aids created by an individual to provide internal cues to prior knowledge, are also seen to be externally and internally oriented. This aspect of Vygotsky's work was taken further by Newman, Griffin, and Cole (1989), who suggested that physical manipulative tools have a role to play in the zone of proximal development. They argued that a child's understanding of the function of a tool comes from involvement in culturally organized activities in which the tool is involved, rather than from unaided exploration of the tool itself. Teachers provide students with a special set of manipulative tools in science lessons, including microscopes and measuring instruments. However, little is known about how the use of these tools directly contributes to students' understanding of scientific concepts.

Although Vygotsky's theories have become increasingly popular, there is little research yet to support or refute the applications of these theories into teaching practice. Shayer (2002) argued that if one wants to see the specifics of what is present in various children's ZPDs, and if evidence is required that partially achieved schemes are in children's minds before they achieve them, it is in the work of Piaget, rather than Vygotsky, that this evidence will be found recorded in fascinating detail.

Feuerstein's Instrumental Enrichment Program

The purpose of the Instrumental Enrichment (IE) Program (Feuerstein, Rand, Hoffman, & Miller, 1980) was to develop learning potential in socioculturally disadvantaged adolescents, many of whom were members of ethnic minority groups. The authors argued that the low level of scholastic achievement and general cognitive performance of these students was due to the insufficient development or inefficient use of the cognitive functions that are pre-requisites of thinking and learning. Inadequate mediated learning experiences were seen as the source of these cognitive deficiencies. The IE program was therefore designed as a remedial enrichment program to correct the deficient cognitive functions. The IE materials included booklets of paper-and-pencil tasks that covered areas such as analytical perception, comparisons, categorization, orientation in space and time, and syllogisms. These booklets were called *instruments*, because their purpose was to 'repair' a number of deficient functions (Kozulin, 2003). The IE materials were content-free, and incorporated a number of graphic devices, such as schematic representations, tables and charts. Activities included coding and decoding, use of models and formulae, different modes of the same problem, generalization and classification. The emphasis on symbolic psychological tools, and process rather than content, was intended to enhance the development of metacognitive awareness and higher order thinking skills.

The IE program appears to be an effective method for the remediation of deficient cognitive functions in culturally different, educationally deprived, and learning-disabled students (Ben-Hur, 1994; Kozulin, 2000). There have been several other experimental tests of Feuerstein's work. Disadvantaged or 'culturally deprived' children, who were exposed to a higher level of the teaching of matrix problems, performed better on Raven's Standard Progressive Matrices test than did children who were given less problems or none (Tzuriel & Feuerstein, 1992). Mediation measures in the mother-child interaction have been reported to predict children's achievement in the first two year's of primary school (Tzuriel, 1997; Klein & Aloni, 1993). An early education program called Bright Start (Haywood, 1986) provided children in Israel with the cognitive tools for school learning. First graders made gains in cognitive performance and achievement tests in mathematics and reading comprehension against test groups over all measures (Tzuriel, Kaniel, Kanner & Haywood, 1997).

Research with retarded persons

Another approach to the understanding of cognitive development is through research with retarded children and adults. The ways in which the cognitive processes of these individuals differ from those of non-retarded people can specify important components, such as the ability to generalize information from one situation to another (Campione & Brown, 1978).

Clinical studies with the mentally retarded are a useful research tool to reveal information about basic developmental processes. Paour has been conducting research since 1968 on the induction of logic structures in mentally retarded persons (Paour, 1992). Paour's 'operatory learning' procedure combines both assessment and a treatment that stimulates cognitive development. At the theoretical level, his research on the learning of logic structures has contributed to the understanding of mental retardation. On a clinical level, it has provided an effective tool for assessment and treatment.

The theory base of CASE

Perusal of the CASE methods and materials shows that the theoretical foundations of CASE can be attributed to the work of Piaget, Vygotsky, and Feuerstein (Shayer 1999, 2002). In the current climate of educational research, the work of Vygotsky is often regarded as politically more palatable than that of Piaget. Shayer (2002), however, argued that the theoretical background and rationale of CASE did not come from just Piaget, nor just from Vygotsky, and certainly not from Vygotsky as an alternative to Piaget.

The idea of exposing young adolescents to cognitive conflict (puzzlement) during CASE lessons in order to encourage construction of the reasoning patterns of formal operations clearly derives from Piaget's notion of disequilibrium, as does the notion of concrete structuring. The content of the CASE *Thinking Science* lessons is also heavily influenced by Piagetian theory, as it covers control and exclusion of variables, ratio and proportionality, compensation and equilibrium, probability, and correlation, all direct representations of Piaget's schemata for formal operational thinking described in Chapter 17 of GLT (Inhelder & Piaget, 1958).

Several aspects of the CASE class-management strategies might well facilitate a student's growth within Vygotsky's zone of proximal development. The use of physical tools is evident in the practical component of the many CASE lessons that use demonstrations to engage students with the problem. Following this aspect of concrete preparation, students are encouraged to participate in discussion with peers, as well as with the teacher, before reflecting on their own thinking. There are opportunities for both the expert/novice interactions of cognitive apprenticeship, as well as the mutual exchanges of participatory approbation, in this metacognitive component of the CASE lessons. Another component of CASE lessons is bridging, that Shayer (1999) attributed to Vygotsky and Feuerstein. CASE teachers use bridging to explicitly encourage students to transfer ideas across the curriculum, so helping them to connect the new experiences with their prior knowledge. Indeed, the CASE strategies appear to provide students with comprehensive mental tool kits with which to enhance their cognitive development.

Table 2.1 shows how the theory bases of Piaget, Vygotsky, and Feuerstein provide the five pillars of CASE: concrete preparation, cognitive conflict, construction and metacognition, and bridging.

Table 2.1
Theory base of the CASE project

THIS TABLE HAS BEEN REMOVED DUE TO
COPYRIGHT RESTRICTIONS

Shayer (1999) acknowledged that although the only theory base mentioned explicitly in the *Thinking Science* materials was that of Piaget, the Vygotskian aspect was implicit in the class management strategy of the sample lessons generated for CASE I in 1980-83, as this was taken from the practice of Feuerstein's Instrumental Enrichment (Shayer & Beasley, 1987). It was only during CASE III (1989-91) and in the writing of *Really Raising Standards* (Adey and Shayer, 1994) that the Vygotskian source was acknowledged explicitly. Shayer attributed this to "driving by the seat of one's pants' at the frontier of knowledge and only later finding out how one did it" (Shayer, 1999). In

Really Raising Standards (Chapter 6), Adey and Shayer (1994) suggested that the Vygotskian aspect was perhaps the engine of CASE and the Piagetian aspect the gearing. The gearing was seen to be involved with getting the right match between the cognitive demand of the material and the cognitive level of the learners.

Shayer (1999) summarised the theoretical background of CASE. From Piaget comes the selection of the different science contexts, that include at least eight of Piaget's formal operational reasoning patterns (schemata). The idea of provoking, and then resolving, cognitive conflict in order to assist students in constructing more powerful personal strategies also stems from Piaget. However, it is from Vygotsky that we gain the speculation that self-construction is only a small part of cognitive development. A greater proportion of a child's cognitive development can be attributed to the internalization of successful performance as seen in another person in their social environment and/or by working collaboratively with more advanced peers in the zone of proximal development. Finally, from Feuerstein (Feuerstein *et al.*, 1980) comes the idea of bridging - taking a term or concept from the context it is learned and applying it in a different but relevant situation. This takes place when teachers learn to transfer the class-management strategies, that characterise CASE lessons, to the rest of their science teaching. Shayer (1999) suggested that it is this bridging by teachers that might be responsible for the long-term, large scale effects of the CASE intervention, rather than the *Thinking Science* lessons themselves.

Jones and Gott (1998) argued for making a link between the CASE program and the procedural knowledge base of science education. They regarded the central premise of CASE - to develop students' ability to learn science through the development of higher order thinking skills - as a sort of 'diet supplement' that is not part of science but renders science more intelligible to students. Their concern was that if CASE is regarded as an add-on element of teaching, there is a danger that it will not become a legitimized part of science curriculum structures. To address this problem, they suggested that the CASE program should be dismembered into procedural content that should become part of the 'diet' itself, a part that is at present missing. Jones and Gott viewed the three essential elements of CASE teaching (cognitive conflict, metacognition and bridging) as ways of teaching that are available across the whole of science to be used at the teacher's discretion. They regarded these essential elements as the CASE 'content', a part of procedural knowledge that can be taught and so increase an individual's ability to cope with problems. In their view, science curricula can be amended to include this procedural knowledge base, with the result that the CASE methodologies would become part of a range of teaching strategies appropriate to a group's age and ability.

Shayer (1999) responded to Jones and Gott. He claimed it is stretching the meaning of the words 'content' and 'knowledge' too far to apply them to the essential elements of CASE such as cognitive conflict, metacognition and bridging. Shayer argued that the aim of CASE is to teach thinking skills, rather than specific science process skills, and that this is fundamentally different from teaching a procedural knowledge base. He also pointed out that teachers gain valuable professional development through teaching CASE. By becoming more aware of the existence of hierarchies of difficulty among the concepts in science, teachers are better able to achieve a match between their planning of lesson sequences and the learning abilities of their students. Shayer also emphasised that CASE professional development provides teachers with class management skills that foster the construction of knowledge by students, so enhancing their cognitive development and school achievement.

Measurement of cognitive development

In many fields there has been a close link between the development of theory and the development of good principles of measurement. However, in psychology this link has not been traditionally strong due to the divergence between quantitative and qualitative research methods (Styles, 1999). These research methods have been seen by some (*e.g.* de Vries, 1974) as mutually exclusive and by others (*e.g.* Andrich & Styles, 1994) as complementary. Michell (1999) traced the history of measurement in psychology, and addressed the issue of whether psychological attributes really are quantitative. He concluded that there are many aspects of human life that are non quantitative, but nevertheless can be investigated in a scientific manner in terms of their categories.

Early measurement practices assumed a connection between intelligence and physical characteristics such as skull circumference and brain volume. Binet and Simon (1980) developed the first individual test of intellectual functioning in response to a need to identify children for special education (Styles, 1999). They developed a psychological scale with a continuum based on children's latent ability, with questions ordered according to difficulty and the age at which children could be expected to answer them. Although the tests of Binet and Simon assumed the development of intelligence with age, too often the results of such tests have been interpreted as the measurement of an innate unalterable ability.

The test of cognitive developmental level used in this study is *Bond's Logical Operations Test* (BLOT), which is a thirty-five-item multiple choice test used to estimate the Piagetian level of children's cognitive development (Bond, 1976, 1995a). The items in the test are drawn directly from the 16 binary operations and the INRC, described in Chapter 17 of GLT (Inhelder and Piaget, 1958), and are categorised in Tables 2.2.

Table 2.2
BLOT content

THIS TABLE HAS BEEN REMOVED DUE TO COPYRIGHT RESTRICTIONS

(After Bond, 1990)

The BLOT was developed as an alternative to the interview technique of Inhelder and Piaget, and it has a number of advantages over the more traditional Piagetian revised clinical method. A multiple choice pencil and paper test is much easier to administer and less time consuming, both for the researcher and for the subject. It can be conveniently administered simultaneously to a large sample of people. Because the results are directly quantifiable, the test also lends itself to statistical analysis and evaluation. The validity of the BLOT has been confirmed by Bond (1976, 1995b). The usefulness of the test has been endorsed by other workers in the field including Christiansson (1983) and Smith and Knight (1992). Bond (1995b) has demonstrated concurrent validity of his BLOT and the *Piagetian Reasoning Task – Pendulum* (NFER, 1979), a post-test routinely used in the CASE project in the UK. Although BLOT and PRTIII differ in format and only partially overlap in their range of age and ability of the subjects, the Pearson correlation coefficient between raw scores on PRTIII and BLOT was $r = 0.75$ (Bond, 1989). The raw total score of BLOT can be used to estimate the Piagetian level of a child (Bond, 1976). Table 2.3 shows the BLOT threshold scores and the corresponding Piagetian levels of cognitive development.

Table 2.3
BLOT thresholds and Piagetian levels (After Bond, 1976)

THIS TABLE HAS BEEN REMOVED DUE TO COPYRIGHT RESTRICTIONS

The BLOT has poor powers of discrimination at the top end of the scale, although it is particularly discriminating over the period where children typically develop formal operational thinking. The problems associated with the "ceiling effect" of BLOT are well acknowledged:

> The ceiling effect on the BLOT continues to be an ongoing problem: The most cognitively developed kids top out on the test. Although that amounts to only 3 out of the 150 tested for this chapter (at age 15), we could reasonably expect that more and more of these students would "hit the ceiling" as we tracked their development over time. This is further complicated to the extent that some Rasch analysis software routinely imputes an ability estimate for those who get a perfect score, whereas other software packages ignore the perfect scores because they do not have enough information to provide an accurate estimate. Clearly, the BLOT needs more difficult items based on Piaget's specifications if we intend to use it to estimate accurately the cognitive development of our more intellectually able teenagers.
>
> Bond & Fox (2001), page 49

That the BLOT discriminates very well across the range of scores for children who are making the transition from concrete to formal operational thinking helps to make BLOT an ideal instrument for use in this study, because the purpose of the STEP intervention was to change children's thinking from concrete to formal operations.

Because the BLOT items are drawn directly from the actual logical operational processes described by Inhelder & Piaget (1958) to model the development from concrete to formal operational thinking, it follows that the BLOT might be more free of retest effects than other types of tests (for example, tests which measure what has been learned, rather than the capacity to learn). Indeed, the report of the original development of the BLOT provides empirical evidence of this property of the test. Bond (1976) reported that a randomly drawn sub-sample (N=91) of the larger broad ability sample (N=899) of secondary school students used in the test development was subjected to a second administration of the BLOT, in order to provide an estimate of its reliability (test-retest stability over time). The analyses of that time did not benefit from Rasch modeling, but Bond reported a correlation of $r = 0.9$ ($p < 0.001$) between scores over a time interval of six weeks. The retest effect seems to have been minimal, about one-quarter of a standard deviation (mean test = 27.34 [SD 6.37]; mean retest = 28.9 [SD 5.81]).

Rasch Analysis

An important characteristic of the present study is the use of the Rasch model, which can help transform raw data from the human sciences into abstract, equal-interval scales. The measurement tool constructed by Rasch analysis allows, for example, test items and student performance in BLOT to be measured on the same true interval scale, so that a meaningful assessment can be made of children's cognitive development. Not many studies of child development using Rasch analysis have been published (a notable exception is the work of Susan Embretson), a feature that adds interest to the present study.

Although measuring change in ability and achievement over time has had both theoretical and pragmatic attraction for those in education and psychology, the field has been fraught with apparently insurmountable difficulties. Early research often used ability indicators with little obvious connection to school learning. Furthermore, data analysis approaches based in classical test theory "led to many seemingly irresolvable conflicts" in that both the mean level of change and the meaning of that change "depend on the distribution of the initial scores, which is population specific" (Embretson, 1991).

The theoretical base of the CASE research has shown that estimates of cognitive development using a Piagetian framework have modest to strong correlations with school achievement indicators (Shayer & Adey, 1994; Endler & Bond, 2001; Stanbridge, 2002; Bond, 2001). Early attempts to conceptualize the measurement of change within the framework of item response theory (e.g. Fisher, 1976) revealed the promising potential of latent trait models to resolve the difficulties alluded to above. In particular, the estimates of the effect parameters in the Fisher model are independent of the distribution of the ability in the sample, as the person parameters can be factored out of the equations for their estimation. Andersen's (1985) Rasch-based model is illustrative for the purposes of the current research, as it was developed to model the repeated administration of the same items over time.

It is well known that the family of Rasch models (Rasch, 1960; Fischer & Molenaar, 1995) has strong measurement properties including that of specific objectivity, *i.e.* item difficulty estimates which are free of person-distribution constraints and person estimates that are correspondingly item-distribution-free. Embretson's (1991) multidimensional Rasch model for learning and change (MRMLC) provided two key improvements over the Andersen model. First, it provided for maximum-likelihood estimations of person ability estimates. Second, it provided for estimates for repeated testing over time, rather than being constrained to repeated administration of the identical test over time. Embretson (1991) argued this as necessary to avoid well known retest effects (incorporating memory, response consistency, and practice effects).

Measurement and analysis procedures adopted for this research project are not nearly as ambitious as those of Andersen (1985) and Embretson (1991) in which all relevant parameters are estimated simultaneously. Because the test of cognitive development was used annually to retest the students in this study, individual person ability estimates were calculated for each occasion with item difficulty estimates anchored at the values derived from its first administration to a large sample (N=633) of Molalla school-children. Furthermore, the stability of BLOT has been confirmed: retest effects were minimal over an interval of six weeks (Bond, 1976).

The appropriateness of applying the Rasch model to Piagetian measures of stages of thinking has been endorsed by Hautämaki (1989). Other applications of Rasch analysis include a test of computer anxiety (King & Bond, 1996) and the validation of written tests to measure Piagetian formal operational ability (Bond, 1995a, b). Rasch analysis has also been used by Bunting (1993) in a study of Piaget's control of variables schema and by Lake (1996) in an investigation into the concept of time held by trainee teachers in Papua New Guinea.

Rasch analysis is regarded as an ideal technique to analyse the results of tests such as BLOT because Rasch models satisfy Thurstone's requirements for fundamental measurement and so provide a suitable link between measurement and theory. Thurstone (1969) described the requirements of fundamental measurement as invariance, unidimensionality, and addivity. He stated that the difficulty of the items in a test should be invariant, *i.e.* independent of the people tested, located on a continuum or scale, and that the location of the items on the scale should be additive.

Wright (1985) regarded the Rasch measurement models as the most important advance in psychometrics of the century. The Danish mathematician Georg Rasch developed these models between 1951 and 1959 and presented his argument and principles in his book *Probabilistic Models for Some Intelligence and Attainment Tests* (1960). The work of Rasch is particularly significant in that it has brought the principles of scientific objectivity to the highly subjective field of educational psychology (Andrich, 1988; Bond & Fox, 2001).

When an item bank, rather than a test, has been constructed to fit the Rasch model, the result is a scientific tool that is both simple to use and very effective. Rasch analysis allows for the items to be calibrated and for persons to be measured on a common interval scale. As such it is appropriate for the construction of rating scale models and for the analysis of graded performance data. It has been used for the validation of educational test items and for the subsequent measurement of individual achievement levels. Once such an item bank has been calibrated, a person's attainment can be measured with a sequence of items evenly spaced over the region where the person is thought to be. Because the items are of graded difficulty, students can choose which questions to answer and work until they reach their limit. The analysis of the response pattern allows each student's item responses to be diagnosed in detail (Wright, 1979).

The Rasch model makes two major assertions about the probability of an individual correctly answering an item in a test situation (Bond & Fox, 2001). The first assertion is that the person's ability remains the same (at a given time), irrespective of the particular test items taken. The second is that the difficulty of the item is the same, regardless of whomever attempts it. The intermediate zone seen in Figure 2.2 represents students responding to items on which they have a 50% chance of gaining a correct answer. It is therefore possible to estimate the ability of a candidate in terms of which items lie within this zone of challenge, rather than in terms of the impossibility of items. Item difficulty can therefore be used to define candidate ability.

The Rasch analysis model asserts that the probability of a correct response is a function of the difference between the person's ability and the difficulty of an item. Both item difficulty and person ability are estimated using a procedure known as maximum likelihood estimation. Using a mathematical model, predictions are made as to the most likely outcome (1 or 0), so producing a predicted score set that is then compared with the actual observed score set. If there is a discrepancy between the predicted and observed scores that is greater than a stringently preset level, then revision of the estimated item difficulties and person abilities occurs. The procedure stops when the differences are minimised (Bond & Fox, 2001).

THIS TABLE HAS BEEN REMOVED DUE TO
COPYRIGHT RESTRICTIONS

(after Bond, 1997)

Figure 2.2 Schematic diagram of the patterns that emerge from a Rasch analysis

Person estimates, also known as ability estimates, are more reliable than raw scores as they take into account the particular responses of the candidates to all the test items, as well as more sensitively estimating the differences in ability. Ability estimates are expressed in units of measurement known as logits (log odds units). The higher the value of the unit, the more able is the candidate. Candidates with a positive ability estimate will find items of average difficulty within their grasp. Negative values indicate that the candidate will find items of average difficulty to be challenging or overwhelming (McNamara, 1996).

Estimates of item difficulty can also be obtained from Rasch analysis. The same logit scale can be used for estimates of both person ability and difficulty of item, because these are expressed in terms of each other. If estimates of item difficulty and person ability are represented together graphically, an item-person map (or Wright map) is produced (see Figure 5.2 in Chapter 5). The ability estimate of each of the candidates is plotted as **X**. Item numbers are used to plot item difficulty estimates on the right. The top of the scale represents greater item difficulty and greater person ability. The item-person map shows the relationship between the items and the candidates.

The software package for Rasch analysis used in this study was *Quest - The Interactive Test Analysis System* (Adams and Khoo, 1993). Quest uses Joint Maximum Likelihood Estimation to provide item estimates, person estimates (ability estimates), reliability indices, and fit statistics in the form of tables and maps. The fit statistics are reported as unstandardised mean squares and standardised *t* statistics. Persons with extreme scores (all correct or incorrect) are not assigned an ability estimate by the 1993 version of Quest. A beta version of the Quest software (Quest 90 PISA version August 1999), an enhanced version of Quest that uses the Warm estimate, was used to obtain an ability estimate for candidates who scored 35/35 in the 1999/2000/2001/20002 BLOT tests.

Criticisms have been made of the Rasch model on the grounds that items validated for difficulty at one point in time may not necessarily retain the same level of difficulty later. Also questionable is the claim that the relative level of difficulty will remain unchanged despite differences in the learning experiences of the students. The nature of the different student groups participating in a test is another variable. It would seem logical that the difficulty of test items is not always independent of such variables as social class, ethnic group or teaching method, and it follows that content and context bias may therefore be a problem in some cases (Gipps, 1994). However, problems such as these should be informative, indicating where the invariance principle is violated and sending the interested and informed researcher on the road to constructing an explanation or revision. Rasch measurement remains the model of choice in this study, and for many other researchers in the social sciences, because it provides a practical and objective approach to data analysis.

CHAPTER 3

A review of the CASE research

This chapter includes a review of the research which informed the original CASE project in Britain, and a summary of other CASE studies conducted in Pakistan, Denmark, Finland, Malawi and Australia.

Research that informed the original CASE project

Much of the research that informed the original CASE project is grounded in Piagetian theory (especially Inhelder & Piaget, 1958). There was strong interest among educators in Piaget's ideas in the 1960s and early 1970s, and several replications were performed of Inhelder's original research on the acquisition of formal operational thinking. Lovell (1961) tested a sample of 200 subjects (eight years to adult) with 10 of the problems described in *The Growth of Logical Thinking* (*GLT*), that generally substantiated the work of Inhelder and Piaget. Jackson (1965) also performed a supportive replication with 48 children of normal intelligence aged five to 15 years and 40 subnormal children aged seven to 15 years. His research demonstrated consistency of achievement across six tasks from *GLT* and was also successful in estimating the proportion of his sample that had reached each Piagetian sub-stage of cognitive development. He found no formal thinkers in the subnormal group. Lovell and Shields (1967) found that of 50 gifted eight- to 10-year-old children, only 10% were at the level of formal operations in spite of their records of giftedness. This particular replication of Inhelder's original research suggested that the distribution of formal operational thinking might be rather more restricted in children than was previously thought.

In the USA, Lawson and Blake (1976) investigated the performance of 68 high school biology students on three of Inhelder's tasks. The students ranged in age from 14 years seven months to 17 years 10 months, with a mean age of 15.5 years. Lawson and Blake concluded that 15% of the group were at the early concrete stage of cognitive development, 42% were mature concrete thinkers, 35% were at the early formal stage and only 8% of the students were operating at the mature formal level. This study raised the concern that there might be fewer formal thinkers than was previously expected among high school students. The implication drawn from the work of Lawson and Blake was that children restricted to concrete operational thinking are generally unable to develop formal concepts from standard lecture type high school classes and that they need concrete examples to work with before reflective abstraction can take place. In a later study with 507 students aged between 11.5 and 20 years, Lawson, Karplus and Adi (1978) administered a paper and pencil test of ability in proportions, probability, correlations and propositional logic. He concluded that there was an increase with age for performance on items requiring the formal schemata, but that this was not as clearly demonstrated for what he described as propositional logic items.

A number of studies have been carried out to investigate the distribution of Piagetian stages in populations of school children. In the most comprehensive of these, Shayer, Küchemann and Wylam (1976) calculated the proportion of children showing early and mature concrete operational thinking and early and mature formal operational thinking in a representative sample of more than 10,000 children between the ages of nine and 14 in Great Britain. They used demonstrated class-tasks derived from Inhelder's individual interview situations. Shayer *et al* (1976) concluded that most children in early adolescence showed rapid development in concrete thinking, but only one fifth showed the further development of formal operational thought. In a further study of 1,200 15- to 16-year-olds, Shayer and Wylam (1978) found no increase in the proportion of students showing formal operational thinking beyond the age of 15 years. There was no increase for girls after the age of 14, whereas the boys continued to develop for an additional year.

Although Inhelder and Piaget gave some general indication of the ages at which their sample of children was capable of formal operational thought (see Bond, 2001, p. 75 Table 4.1), it would not be reasonable to expect to replicate their findings in populations of children from vastly different settings. Indeed, Piaget (1972) later claimed that their research had been conducted with a somewhat privileged population and that generalisations to all subjects cannot be made from their conclusions. The results of Shayer, Küchemann and Wylam (1976) were from children in comprehensive schools, selected to obtain a strictly representative sample of the British population. They also showed that, if a selective school population is sampled by testing children from grammar schools and private schools only, formal operational thinking is detected from the beginning of secondary schooling (more in line with the general view derived from *GLT*).

Many attempts to accelerate cognitive development have been reported in the literature. Most of the earlier studies were rather limited in scope, consisting of short term projects, rarely extending beyond two months at the rate of one intervention per week. Siegler, Liebert and Liebert (1973) taught 10-year-old children to use the control of variables schema to solve Inhelder's pendulum problem. As with many of the earlier studies, there was no attempt to test for the transfer of this skill to the context of other tasks. Lawson and Wollman (1976) succeeded in teaching 10- to 12-year-olds to control variables, and went further to demonstrate the specific transfer of that skill to novel tasks that also involved the control of variables. In 1982 Lawson and Snitgen showed that college freshmen made significant increases in formal reasoning ability after exposure to a special one semester biology program, but no evidence of generalised transfer to new contexts was demonstrated.

A few short term intervention studies support the notion that generalised transfer can occur from a cognitive acceleration program. Campione and Brown (1974) investigated transfer of pre-training second grade children in dimension-abstracted oddity tasks. They concluded that the probability of transfer from one task to another depends on the similarity of the task formats. Kuhn and Angelev (1976) gave 8- to 11-year-olds a problem solving exercise, where the children were required to explain their strategies by thinking aloud. The experimental group performed better than a control group on both a pendulum and a chemical combinations post-test, although the intervention exercise was based on a control of variables task. Similarly, Rosenthal (1979) found that general transfer took place in 15-year-old girls after two one-hour training sessions that were designed to equip them with the cognitive strategies for solving control of variables problems. Klauer (1989) found transfer to new tasks requiring inductive thinking in kindergarten children, who had been taught inductive reasoning as training for analogical transfer. Adey and Shayer (1990) suggested that interventions which result in general transfer provide students with the essential mental tools to enable them to construct the formal schemata for themselves. They hypothesised that it is the process of students constructing their own meanings that leads to the cognitive restructuring required for transition to the next stage of development.

Shayer (1999) distinguished between context-independent and context-delivered interventions. A context-independent intervention, such as Feuerstein's Instrumental Enrichment, is delivered as a discrete course, quite separate from the school curriculum. On the other hand, a context-delivered intervention, such as CASE, is situated in the context of a particular school subject. Shayer reported that whereas both types of intervention bring about gains on psychological tests, only the context-delivered programs were accompanied by long-term effects on student achievement. The success of CASE in promoting the acceleration of formal thinking in school children in the UK has been clearly demonstrated (Adey & Shayer, 1994).

A history of CASE in the United Kingdom

Cognitive Acceleration through Science Education (CASE) was originally funded as a research project by the British Social Science Research Council. The project began at Chelsea College, London, with a pilot study known as CASE I (1981-1983). This was followed by the main project, that included an evaluation of the intervention, CASE II (1984-1987). Six years of applied research under the direction of Michael Shayer and Philip Adey, with support from Carolyn Yates, were needed to design, generate and refine the methodology. In CASE III (1989-1991) the focus was on a detailed description of the teaching skills involved in the intervention program. CASE is now being taught at hundreds of schools throughout the UK, by teachers who have received training in the CASE classroom methods from the professional development programs based at King’s College, London.

The results of the CASE II research showed that the two-year intervention had a substantial effect on the cognitive development on students, as well as a long-term effect on their school achievement (Adey & Shayer, 1990, 1991, 1994; Shayer & Adey, 1992a, 1992b, 1993). Shayer (1993) reported effect sizes between 0.67 and 1.12 standard deviations per school on the cognitive post-tests. A 30-percentile difference was found between the pre/post-test changes for students in eight CASE schools and norms for 14, 000 British school children aged 10-16 years (Shayer, Küchemann, & Wylam, 1976; Shayer & Wylam, 1978). A long term effect was seen in science (three years after the intervention), as well as far transfer effects, away from the context of delivery, into mathematics and English. Effect sizes were between 0.3 and 1.0 standard deviations, which correspond with a half, and one GCSE grade (Adey & Shayer, 2002).

The CASE II results were based on relatively small scale data from around 130 students in each of the experimental and control groups (Adey & Shayer, 2002). More convincing evidence of the effect of the CASE strategies on cognitive development and school achievement came from the data subsequently collected from over two thousand pupils from 11 schools whose teachers participated in the 1994-1996 CASE Professional Development programs. The students in the CASE schools scored consistently higher grades than their peers in the control schools. The average gain in the GCSE for the CASE schools was 1.05 grades (0.6 standard deviations) in science, 0.95 grades (0.5 standard deviations) in mathematics, and 0.90 grades (0.57 standard deviations) in English (Adey & Shayer, 2002). The improved student achievement in subjects other than science has been attributed to the CASE intervention having an effect on general intellectual growth, rather on domain-specific skills (Adey & Shayer, 1994, 2002).

Adey (1992) reported gender differences in the long term effects of the cognitive acceleration program. The proportion of girls reaching formal operational thinking did not increase after the age of 14 years, whereas this effect was seen later, at 15 years, in boys.

Further, the age at which the program was started was found to be relevant. Boys who began CASE at age 12+ years performed significantly better than controls in science and mathematics in the GCSE, whereas 12+ girls performed better than controls in English. The boys who began at age 11 years, unlike their female peers, showed no effects of the intervention in the GCSE. Adey (1992) further showed that boys exposed to the intervention course achieved an average of 40% more grades of C or above in the science GCSE than did controls, whereas the gain for girls was lower, at 25%. This work raises the idea that there might be different critical periods of learning for boys and girls and that gender might also be related to performance in certain subjects: an important point, given some of the issues currently being raised about boys in education. High achievement gains were reported for the full ability range of pre-intervention students (Adey and Shayer, 1994). They concluded that the students who made the greatest gains in academic achievement were not restricted to those who had more ground to make up, nor to those who were more able and therefore perhaps more ready to move on to formal thinking.

Leo and Galloway (1996) argued that some of the results of CASE should be attributed to student motivation. They referred to research suggesting that children who have low self-concepts tend to lack metacognitive skills (Carr, Borkowski, & Maxwell, 1991) and offered this as an explanation for why some students might not respond to the CASE strategies. McLellan and Adey (1999) raised the question of whether the success of CASE could be attributed, at least in part, to the fact that students of schools that opt for the program already display a higher motivational profile than students in other schools. They reported that the early results of a long term study of the effects of CASE on achievement and motivation suggested that a higher proportion of students in CASE schools exhibit what Dweck and Leggett (1988) term adaptive motivational patterns, that might make students more receptive to any intervention they receive. It remains to be seen whether it is only “good schools” that opt into successful projects such as CASE.

There have been several CASE studies in other countries. These studies are compared in Table 3.1 and are described individually later in this chapter. Cognitive acceleration projects in the UK now exist not only in science (CASE), but also in mathematics (CAME), technology (CATE) and other subjects. Some of the latest CASE research targets primary school children and suggests that a differential between children’s potential and their performance might be detected as early as 5 years (Adey & Shayer, 2002).

Although little doubt remains that the CASE strategies do enhance cognitive development and student achievement, there is still much more to be learned about the changes in cognitive processing that actually takes place. For example, it is not at all clear how the CASE strategies assist students in making the transition from concrete to formal operations. Now that detailed analysis of videotapes has provided empirical evidence of how children’s conceptual understanding evolves during physics tasks (von Aufschnaiter, 2003), it might well be that similar close observations of students performing CASE activities could shed further light on how CASE really does raise standards.

CASE in Pakistan

Iqbal and Shayer (1995) reported that the level of cognitive demand of the Pakistan science curriculum for 11 to 13 year olds was far in excess of the level of thinking of those students as revealed by Piagetian tests. This research lead to the introduction of the CASE intervention in three schools with the goal of improving the match between curriculum

Table 3.1 A comparison of various CASE studies

	UK	Pakistan	Denmark	Finland	Malawi	Australia	USA (STEP)
# of CASE activities	30	30	15	28	30	25-30	13-21
Time period (years)	2	2	1	1	2-3	2+	3 or 4
Age of students (yrs)	11+ , 12+	11+ , 12+	16+	12+	13 - 22	13+	11+, 12+, 13+
Cognitive test	Piagetian Reasoning Tasks	Piagetian Reasoning Tasks	Modified Piagetian Reasoning Task	Piagetian Reasoning Tasks	Piagetian Reasoning Tasks	*Bond's Logical Operations Test*	*Bond's Logical Operations Test*
Achievement test	General Cert. of Secondary Education (science, mathematics and English)	School science examinations	No data available	School GPA	Malawi School Cert. of Education (science, mathematics, & English)	Australian Schools Science Competition, Queensland Core Skills Test	State-mandated tests (mathematics, reading & literature)
Teacher preparation	Extensive professional development	Meetings with CASE experts	2 days of training	No formal training	Eight workshops, regular meetings	No formal training	4 & 2 day workshops

demand and students' ability.

Two fee-paying schools and one government school took part in the project and teachers attended professional development workshops in the CASE methods at the University of Punjab. Following the practice used in the United Kingdom, the intervention took place over a two-year period. A separate class of students who did not receive the intervention acted as a control group in each school. Pre- and post- tests of cognitive development were conducted using *Piagetian Reasoning Tasks* (NFER, 1979). Two post-tests were carried out, one six months after the start of the intervention and the second a year after the intervention ended. The end of year examinations developed and administered by each school were used as the measures of student achievement. As these examinations were specific to each school, comparisons were made between experimental and control classes within schools, rather than between schools.

Iqbal and Shayer (2000) reported that the proportion of students who showed evidence of formal operational thinking was higher in the students who had experienced the CASE intervention. The mean cognitive development level for the CASE students was found to be at the early formal stage (IIIA), whereas the mean for the control groups was at the early concrete level (IA), even though the control groups in each school were initially superior to the experimental groups. Iqbal and Shayer (2000) found that the effect of the intervention was to change the mean of the experimental group on the pretest by about 30 percentile points in relation to British norms, an almost exact replication of the results in Britain (Adey and Shayer, 1991). They also reported significant gains with effect sizes between 0.47 and 1.18 standard deviations on the delayed post-tests. Iqbal and Shayer suggested that girls might have gained benefit from the student discussion in CASE lessons, student discussion not being a common feature of science classes in Pakistan. Bridging, in which teachers helped students to connect the CASE ideas with other contexts, was thought to have been crucial to the success of the program.

CASE in Denmark

Denmark has nine years of compulsory education (the "folk school", 7-16 years of age), after which students have a number of options for continuing their education. About 20% of the students aged 16+ years choose the mathematics line of a general upper secondary education, that lasts a further three years (the "gymnasium"). The transition from folk school to gymnasium is a difficult time for many students, because they need to adjust to changes in pedagogy and increased curriculum demand (Krogh & Thomsen, 2001). Furthermore, many students lack the higher order thinking skills necessary for understanding subjects such as physics on the level at which they are taught in the gymnasium (Thomsen, personal communication 2002).

In response to this problem of poorly developed thinking skills, Poul Thomsen and Jens Holbech (Centre for Studies in Science Education, University of Aarhus) set up the Higher Order Thinking in Physics Education (HOT-Physics) project. Twenty-two teachers were given two half-day training sessions, in which the CASE materials and methods were introduced. Before the teaching started in August 2000, the Science Reasoning Task III 'pendulum' was used to estimate the Piagetian level of the students. The pilot phase comprised 22 classes with about 600 students, and a further 15 control classes from the same schools were also given the pre-test. HOT-Physics lessons were then delivered in the autumn of 2000 to the first year students of upper secondary school. Owing to time constraints, only particular CASE activities were taught (control of variables, compensation, compound variables and equilibrium). Most classes were tested again with the 'pendulum' task at the end of the school year.

There were no significant differences between the HOT-physics and control groups at the beginning of the project, however at the end of the year the experimental classes were 1.3 standard deviations above the controls. The control classes had an average gain of half a Piagetian level, whereas the gain for the HOT-physics students was nearly a whole level (Thomsen, personal communication 2002).

CASE in Finland

A double intervention study involving both CASE and its mathematics counterpart CAME (Adhami, Johnson & Shayer, 1998) has taken place in Finland (Kuusela & Hautamäki, 2002). The aim of the study was to evaluate and compare the CASE and CAME interventions in new contexts. The subjects were 6th grade pupils (N = 276, average age 12 years) in the 14 primary schools of Vihti, which has a population of 23,000. The design was a randomised model, with pre- and post-testing in 6th grade, followed by further post-tests in 7th and 9th grades. The teaching was carried out in 28 double lessons in the 1997-98 school year. The measurement scales for CASE were the Learning-to-Learn Assessment Scales (Hautamäki, Arinen, Bergholm, Hautamäki, Kupiainen, Kuusela, Lehto, Niemivirta, & Scheinin, 1999) and Science Reasoning Tasks Volume and Heaviness and Pendulum (NFER, 1979).

At the start, the cognitive competence of the students was at an average level for Finland (Kuusela & Hautamäki, 2002). The results of the study showed statistically significant and lasting effect sizes for both CASE and CAME. The proportion of formal thinkers in the CASE students was substantially increased in relation to the control. Gains were higher for boys initially, but higher for girls by the third test. A surprising outcome was that positive effects were also seen in the control students, who studied with the experimental students in all lessons except those of the intervention.

CASE in Malawi

The effect of CASE on the performance of secondary school pupils has also been investigated in Malawi (Mbano, 2003). The aim of the study was to discover whether evidence could be found for the critical period for cognitive transition from concrete to formal operations at 12 - 14 years of age.

The sample population of 694 children included control and experimental groups of boys and girls from seven schools. The pupils in the experimental schools were taught CASE lessons at the rate of one lesson per fortnight. The teachers attended eight training workshops over a period of three years and were visited in schools at least twice a term. The Science Reasoning Tasks SRT II (Volume and Heaviness) and SRT III (The pendulum) were the pre- and post-tests used to assess cognitive development. The Malawi School Certificate of Education results in biology, physical science, mathematics and English of the sample population were analysed in order to explore whether the CASE intervention improved the academic performance of pupils.

The older pupils, aged 16-17 years, in Malawi made similar cognitive gains to those reported for younger pupils in the British CASE studies. The effect size for girls was 0.9 standard deviations and for boys 1.3 standard deviations. The author interpreted these results as evidence that the critical period for cognitive transition from concrete to formal operations at 12 - 14 years of age does not exist. She also reported that girls, who were on average a year younger than boys in the same class, and older boys, had lower academic achievement than younger boys. The results of interviews revealed that CASE was seen to be beneficial by both teachers and pupils.

CASE in Australia: **recent relevant research by the author**

In the Australian CASE study undertaken by the current author and her Masters supervisor, observations were made of the progressive change in the cognitive development and student achievement of 141 students over the course of their secondary education in a private school in North Queensland (Endler, 1998 a, b; Endler & Bond, 2001, 2002). Cognitive development of the students was measured in years 8 (mean age 13+), 10 (15+) and 12 (17+) using *Bond's Logical Operations Test* (BLOT). Rasch analysis of each of the data sets provided ability estimates for students.

Only 29 students from the group that began the intervention in 1993 remained in the school throughout the five-year period to be tested on all three occasions. This rate of attrition (45% of the original sample was replaced by new students) was typical for this school, situated in a garrison city with a transient population. Data from these 29 students were analysed in order to investigate the children's cognitive development across the five year period. Increased cognitive growth took place between age 13+ and 17+ years for most CASE students (t[df], p = 7.93[28], <0.01), with the greatest change occurring between age 13+ and 15+ (t[df], p = 8.34[28], <0.001). These were cognitive gains of large effect size (0.8 SDs).

The cognitive ability estimates of each of the students who had been tested with BLOT on all occasions were analysed in order to investigate the cognitive development of individuals over time. When the progress of individual students was observed, remarkable heterogeneity was evident within the group. Although 16 out of the 29 students revealed the general group pattern of faster mental growth in the period between age 13+ to 15+, this was certainly not the case for all. A few students in the sample were identified as late developers (growth spurt between age 15+ and 17+), whilst others appeared to have remained in a more or less steady cognitive state throughout the five years of secondary school.

Students who showed the greatest growth in the Australian study came from a variety of starting levels of cognitive development. Adey and Shayer (1994) also reported high gains from the full range of pre-intervention students in the UK. In both studies, the students who made the greatest gains in academic achievement were not restricted to those who had much ground to make up, nor to those who were more able, and therefore perhaps more ready to move on to formal thinking.

The 29 students who were tested three times over the period of this study comprised 16 girls and 13 boys. The pattern of cognitive development shown by the group overall was reflected by that of the boys, a growth spurt between age 13+ and 15+ followed by a plateau from age 15+. Although the mean cognitive level of boys exceeded that of girls at age 13+ and 15+, further analysis showed that there was no significant difference between the rate of change of ability estimates between the sexes when boys and girls were analysed separately for the first two-year period. Both sexes showed strongly significant increases in mental growth between mean age 13+ and 15+ (males: t[df], p = 4.51[12], <0.001; females: t[df], p = 4.78[15], <0.001) . However, girls showed a further significant increase from age 15+ to 17+ (t[df], p = 2.42[15], <0.05), whereas the change for boys over the last two years of secondary school was not significant. In contrast, the girls' cognitive development showed a growth spurt from age 13+ to 15+, and continued to develop for the next two years.

These results, which revealed little early difference between the sexes, contrast with the findings reported by Adey (1992). The pattern of cognitive development for the Australian boys appeared to correspond more closely with that for British children, as an increase was seen up to the age of 15 followed by a plateau (Shayer and Adey, 1981). However, unlike their British peers, the small group of girls in our Australian study showed gains in cognitive development beyond the age of 14 years. The achievement of girls in science and mathematics had been actively promoted in this school for several years. Gender independent texts and assessment materials are in use in Queensland, and teachers are routinely prompted to guard against gender bias in the classroom. The results of the study suggest that the promotion of these girls' science achievement might have been successful.

This study further demonstrated that the Shayer and Adey claim of a relationship between cognitive development and student achievement could be supported when using a different measure of cognitive development (BLOT) as well as different indicators of school achievement (the Australian Schools Science Competition at age 15+ and the Queensland Core Skills Test at age 17+) . We found a significant correlation between cognitive development and student achievement for students tested at age 13+, 15+, and 17+. A strong relationship (p = <0.01) was demonstrated between the results of the tests of cognitive development and publicly accredited scores of student achievement, such as results in the national Australian Schools Science Competition, and the students' Queensland Core Skills (QCS) test results and Overall Position (that are derived from State wide tertiary entrance tests taken at age 17+). A high correlation might be predicted between the Australian Schools Science Competition and the BLOT results, because both are reasoning tasks set in a multiple choice format. However, the Queensland Core Skills test is more general in content, incorporating material from a variety of subject areas in the Queensland secondary school curriculum. The Overall Position is calculated from data obtained from both the QCS test and students' school-based results at the end of year 12 and similarly might not be expected to correlate so directly with the BLOT results. Nevertheless, the results suggested that variables related to cognitive development are implicated in student achievement in this particular secondary school.

Students who experienced the CASE 'Thinking Science' course were found to be more cognitively developed than students who joined the school after the intervention had taken place. Unexpected positive effects were also found in these latecomers to the school who became peers of the CASE students. At the time, these gains by students who had not experienced the intervention were attributed to the ethos of the school. However, given that positive effects were also seen in the Finnish control students, an alternative hypothesis could be that the newcomers gained benefit from sharing a learning environment with their more advanced peers who had experienced the CASE strategies.

CHAPTER 4

Method

The purpose of this study was to investigate whether the STEP had enhanced the cognitive development and school achievement of the Molalla River students who experienced the intervention. Cognitive growth was estimated using *Bond's Logical Operations Test*, and student achievement from state-mandated Mathematics and Reading & Literature tests. Pre-intervention performance profiles and student samples from two non-project school districts served as controls. Various statistical techniques were then applied to the data.

Background information

There are several differences between the STEP and the other CASE studies that make the current study unique and of particular educational value. The most important characteristic of the STEP is that the intervention was delivered in a field setting in which teachers received very little professional development or support. Their preparation consisted of short seminars in 1999 and 2000 conducted by Dr. Trevor Bond of James Cook University, Australia. There was no on-going professional development, although the STEP teachers met regularly for problem-solving sessions during the first year. Teachers who joined the schools after 2000 received no formal training in the CASE strategies. Furthermore, fewer CASE lessons were taught in the STEP than in other CASE programs and the period of delivery was over three years, rather than the usual two. The measures of both cognitive development and school achievement were also different from those of the British CASE research. The results of the STEP show what can be achieved when the CASE strategies are delivered in a sub-optimal field setting.

The STEP commenced in September 1999 in the Molalla River School District in rural Oregon in the Northwestern USA. The project was the result of an analysis of the professional development needs of its teachers, performed by the Molalla River School District, taking into account results from the Oregon state-wide student achievement tests and cumulative student failure rates at 6th-10th grade. This analysis was then examined in the light of recent Oregon Department of Education research about student competence in scientific inquiry to identify the focus of the STEP project (Pellens, personal communication 2002).

The Molalla River School District obtained funding for the STEP project from an Oregon Goals 2000 Year-Five Grant, in which the project was presented as an action research based approach to staff development. The general goal of the project was to enhance the teachers' understanding of the relationships between student abilities and the processes of scientific inquiry. Other stated objectives were for teachers to better align the science curriculum with the Oregon content requirements, to learn how to use the Oregon Inquiry Scoring Guide, and to allow for opportunities for students to produce inquiry work samples in science that would better meet the Oregon Department of Education standards.

The overall goals of the STEP project, as articulated in the grant proposal, were that STEP should increase the passing rate of Molalla River students in the state-wide multiple choice achievement test in science at 8th and 10th grade, as well as improving the performance standard of inquiry work samples in all grades. A further goal relating to professional development was to change the instructional delivery used by teachers from structured, to guided, and then to student-initiated inquiry. In a District initiated survey conducted in 1998, the Molalla science teachers had requested training in assessment, research-based science instructional programs, and teaching strategies.

The Director of Instruction of the Molalla River School District proposed that the best way to implement the goals of the STEP would be to use the methodology and instructional materials of the CASE program, given its success in improving student achievement in the United Kingdom. The action plan described in the grant proposal included four days of training in the CASE methods for the teachers in June 1999, followed by a further workshop in January 2000, both conducted by Dr. Trevor Bond of James Cook University, who would also be available for monthly question and answer sessions via email during the school year. It was planned that teachers would meet in monthly study groups to discuss the implementation of the program, become familiar with the Oregon Inquiry Scoring Guide, and develop exemplary scored student work samples.

Participants and setting

Molalla is located in the North Willamette Valley, in the county of Clackamas, Oregon, USA. The population size of the city is 4,920 and of the county 317,700 (Molalla Community Profile 2001). Molalla has a friendly small town atmosphere, and the community logo "Riding high for America" refers to the Molalla Buckaroo, a four-day rodeo centered around the 4th of July holiday. The principal industries of Clackamas county are paper, lumber and agriculture (speciality products, berries and vegetables). The largest employer in the city is the Molalla River School District with 319 employees. The nearest work place training opportunities are at the Clackamas Community College (10 miles), Portland Community College (21 miles) and Portland State University (27 miles).

Molalla has one public library. The percentage of adults (25+ years) who have tertiary qualifications in Molalla is significantly lower than state and national levels of educational attainment, according to the 2000 US census (see Table 4.1).

Table 4.1
Educational attainment of the adultpopulation of Molalla, OR

THIS TABLE HAS BEEN REMOVED DUE TO COPYRIGHT RESTRICTIONS

(from US census 2000)

The Molalla River School District has six elementary schools (1st-5th grades, average age 7-11 years), one middle school (6th-8th grades, 12-14 years), one high school (9th-12th grades, 15-18 years) and one small private elementary school. The Middle School had an enrollment of 645 in December 1998, the High School 702 and the private school 83. The learning environment of the Molalla River Middle and High Schools is typical of the rural North West of the USA. Instructional strategies are traditional, and teaching is closely aligned with the Oregon State Content Standards.

Teachers in the Molalla River Middle School and High School began using the CASE materials in September 1999 with 6th to 10th grades. Following a review of progress at the end of the first year, the teachers agreed to confine the STEP to 7th, 8th and 9th grade classes. The program continued to be delivered to 7th, 8th and 9th grades only from September 2000 until the end of the 2002-3 school year.

Students in 6th to 10th grade were pretested in September 1999 with the BLOT (see Table 4.2). Fortuitously, a small number of 11th grade students were included in the initial testing when they were taking Biology as an elective with a 10th grade class (without this unplanned sample there would have been no cognitive developmental data collected for 11th grade students). The decision was taken to retain this small sample of 11th grade students in the data set because they appeared to be representative of their grade, and their mean cognitive level was similar that of their peers in a cross-sectional profile of a large sample of British children (Shayer & Wylam, 1978). A third factor to influence this decision was that the only comparison to be made with this small group would be Cohort C in 2002.

Table 4.2
The BLOT testing schedule

Time from start of STEP (in months)	0	8	20	32
Month/year	09/1999	05/2000	05/2001	05/2002
	Number of students tested			
6th grade	52	108		
7th grade	181	175	147	
8th grade	148	84	158	77
9th grade	175	77	132	178
10th grade	90			55
11th grade	12			

Students in the 6th to 9th grade, who continued to participate in the STEP project, were retested with BLOT in May 2000, together with newcomers who would have experienced only part of the STEP intervention after joining the school. Students in 7th to 9th grades were again tested in May 2001 and those in 8th to 10th grade in May 2002. The number of students tested on each occasion varied widely according to whether the teachers involved in the STEP intervention were willing to give up class time for the

BLOT testing. In general, if students were absent on the day of the test they were not given a chance to take it later. Test answer sheets were sometimes misplaced and then sent to me at a later date (or not).

Three cohorts of students (A, B, and C) were identified for analysis, that respectively began the intervention at average age 11+ (6th grade), 12+ (7th grade), and 13+ years (8th grade). As is common in school systems where 'social promotion' is the rule, grade level and mean age are almost interchangeable in this study, because children in Oregon schools are rarely promoted or kept back a year, resulting in a relatively narrow age range at each grade level. The criteria for inclusion the master file were that students must have experienced the STEP program delivered to their grade and been tested in the first year of the intervention and on a minimum of one subsequent occasion. Table 4.3 shows the composition of each cohort.

Two subsets of students were identified who missed part of the STEP intervention. These students are not represented in Table 4.3 because they did not receive the whole intervention. One group (N = 23) did not experience the STEP intervention in 1999-2000, but later became peers of Cohort A. Some of these students, who began the STEP in 7th grade (09/00), were new to the school, whereas others had been in classes that did not participate in the intervention in 6th grade. A second subset (N = 21) of students that experienced only part of the STEP intervention were peers of Cohort B whose 8th grade teachers opted out of the STEP in the 2000-2001 school year. This subset of students was tested with the BLOT in September 1999, May 2000 and then in May 2002. The results of the students who missed a year of the intervention were analyzed separately to provide information about the effect of STEP on cognitive development and student achievement.

Table 4.3
Number of students in Cohorts A, B, and C

		Number of students tested		
		Cohort A	**Cohort B**	**Cohort C**
BLOT	overall	76	105	97
	99-00	39	86	39
	00-01	66	92	46
	01-02	34	96	38
Mathematics	overall	51	101	62
	99-00	45	81	57
	00-01	49	97	60
	01-02	50	100	43
Reading & Literature				
	overall	50	101	65
	99-01	46	84	59
	00-01	50	101	59
	01-02	50	86	51

Intervention

The CASE lessons were delivered by the science teachers for each grade level during the normal scheduled science periods. The STEP teachers decided to cover only 25 of the 30 CASE activities (see Table 4.4). The selection of lessons was based on which activities the teachers felt would best enhance their students' scientific inquiry skills. The CASE strategies were delivered at the rate of five to eight activities per year (see Table 4.5). Activities 1-5, that are used to teach control and exclusion of variables, were regarded by the STEP teachers as the key to improving students' skills in scientific inquiry and consequently were reviewed at the start of each successive year. These five lessons are sequenced so that students learn what is a variable, how two variables can interact, how a fair test is constructed, and the kinds of relationships that can be revealed when data are collected.

The full complement of CASE activities spans several types of reasoning that are characteristic of formal operational thinking: control of variables and exclusion of irrelevant variables; ratio and proportionality; compensation and equilibrium; probability and correlation; and the use of abstract models to explain and predict. STEP students were exposed to only a subset of the activities. Furthermore, the number of CASE activities that students received over the three-year period was not consistent across grades and years (see Table 4.5). Some 8th grade teachers did not participate in the program in 2000-01 and 2001-02 and did not undertake any CASE activities with their classes. Of the two 8th grade teachers who did participate in 2000-2001, one taught three new activities and the other taught eight.

The timing of the delivery of the CASE lessons in the STEP was different from that in the British CASE projects, in which activities are usually delivered over two years, commencing at age 11+ years. The students who began the STEP project in September 1999 were of average age 11+ years (Cohort A), 12+ (Cohort B) and 13+ (Cohort C). Cohorts A and B received a three-year program of CASE activities (see Table 4.5). However students in Cohort C experienced only 13 activities over a two-year period because it was decided to confine the intervention to 7th through 9th grades.

Table 4.4
CASE *activities* (after Adey, Shayer & Yates, 1991)

THIS TABLE HAS BEEN REMOVED DUE TO
COPYRIGHT RESTRICTIONS

Table 4.5
CASE activities taught in the Molalla River School District 1999-2002.

	STEP Cohorts								
	A (6th grade 1999, age 11+)			**B** (7th grade 1999, age 12+)				**C** (8th grade 1999, age 13+)	
CASE activities	**99-00**	**00-01**	**01-02**	**99-00**	**00-01**		**01-02**	**99-00**	**00-01**
1	√	√	√	√	√	√	√	√	√
2	√	√	√	√	√	√	√	√	√
3	√	√	√	√	√	√	√	√	√
4	√	√	√	√	√	√	√	√	√
5	√	√	√	√	√	√	√	√	√
6									
7		√							
8			√		√	√			
9		√							
10			√		√				
11									
12									
13			√		√				
14			√		√	√			
15			√		√	√			
16			√		√				
17									
18									
19			√		√				
20			√		√				
21		√							
22		√							
23							√		√
24							√		√
25							√		√
26							√		√
27							√		√
28							√		√
29							√		√
30							√		√
Total CASE activities	**Cohort A= 17**			**Cohort B: 16 or 21** (depending on the teacher in 8th grade)				**Cohort C= 13**	

The typical CASE lesson plan

In the UK the published CASE materials, known as *Thinking Science* (Adey, Shayer & Yates, 1991), are usually presented to 12- to 14-year-olds in the first two years of their secondary schooling. Although the timing was different for the STEP project, the teachers essentially followed the class-management plan outlined in the CASE teachers' manual.

A typical CASE lesson takes at least one hour to administer and lesson plans from the *Thinking Science* kit include four phases: concrete preparation, work on the task, construction/ metacognition and bridging (see Table 4.6). In the first phase the teacher introduces the students to a problem, such as finding out why some objects float while others sink in Lesson 27 'Floating and Sinking' (see Appendix 1 for the lesson plan and student worksheets).

Table 4.6
The typical CASE lesson plan

THIS TABLE HAS BEEN REMOVED DUE TO
COPYRIGHT RESTRICTIONS

(from Iqbal and Shayer, 2000)

Lesson 27 'Floating and Sinking' tackles the compound variable density through hands-on experiences with buoyancy. In trying to predict whether or not jars will float or sink, students discover that neither volume nor mass alone provide sufficient information for an accurate prediction to be made. Rather, both mass and density must be considered when predicting whether a jar will float or sink.

The teacher prepares the students to engage with the problem by setting it in context and introducing the new vocabulary the students will need to talk about the problem. The strategy used is one of concrete experience - in this case jars of the same volume but different masses are tested in a tank of water. The conclusion is that the heavier the jar, the more likely it is to sink. The teacher then has children repeat the experiment with jars of equal mass but different volumes. The conclusion here is a little more confusing, that the smaller the jar, the more likely it is to sink. Data are recorded in a classification table (Worksheet in Appendix 1).

The teacher now introduces the element of cognitive conflict by providing two new jars, that do not perform according to the children's predictions. For example, jar X is the same volume as jar 3 (that floated), and the same mass as jar C (that floated), but in fact sinks. The sinking of jar X is a discrepant event for many students, providing them with a cognitive jolt. The aim of this cognitive conflict phase of the lesson is to engage students in a struggle with the problem, and to challenge their prior knowledge and personal theories. The teacher plays a vital supportive role by challenging the students to look at the problem from different perspectives and by promoting a discussion of the difficulties experienced. It would typically take about 15 minutes to administer this introduction to the lesson.

Following the introduction, students typically collaborate together in small groups of four or more to work on a task. In the 'Floating and Sinking' lesson, students play a card game (see Workcard A, Appendix 1), that provides a further opportunity for them to discover that both volume and mass have to be considered before an accurate prediction can be made about whether something will float or sink. Twenty to thirty minutes are usually allowed for the group to perform the task and discuss their solutions with each other and with the teacher.

Typically, each group would then present their results to the whole class, with the teacher chairing the discussion. Cross-questioning of students by each other is encouraged. The teacher uses questions such as “How did you solve that?” and “Can you explain why you think that?” Students are prompted to reflect collaboratively on ideas that emerge in the whole class discussion. Worksheets guide the students in constructing their own knowledge and in revising their former theories. Metacognition occurs when children not only become conscious of their own thinking, but develop and practise their explicit thinking processes.

The final stage of bridging takes place when the teacher helps the students to link the new thought processes to other areas of the science curriculum in particular, as well as to their everyday experiences in general. This is often achieved by the teacher inviting students to think of other contexts where the ideas might be useful. In Activity 27 the problem sheet (see Workcard B, Appendix 1) consolidates the ideas of floating, sinking, and compound variables. It also introduces a new compound variable, momentum. The two variables in this case are mass and speed, and the compound variable is the product of these two simple variables, rather than the ratio as is the case with density.

Instruments

The pre- and post-test data collection instruments were different in the STEP project compared with other CASE studies. The pre/post test cognitive development instruments used in the studies in Britain, Pakistan, Denmark, Finland and Malawi were Piagetian Science Reasoning Tasks (NFER, 1979). Whereas, in the Australian CASE study and again in the STEP project, *Bond's Logical Operations Test* (BLOT) was used as the pre/post cognitive test (Bond, 1976). Like the Science Reasoning Tasks, BLOT is directly grounded in Piagetian theory but has an advantage of being in multiple choice format.

Indications of student achievement were obtained from national Certificate of Secondary Education results in Britain and Malawi, school science results in Pakistan and Finland, and the results of state mandated tests in Australia and in the STEP project. The student achievement indicators used for the STEP students are the Rasch-modeled Oregon State Scores obtained from the state standardised Mathematics, Reading & Literature and Science tests. In Oregon, school achievement is reported annually for mathematics and reading & literature, allowing trends over time to be followed for individual STEP students. Statewide testing in science does not take place annually in Oregon, so no data were available to follow science achievement over time.

Student achievement data

In 1996, the Oregon Department of Education introduced standards of student performance and developed assessment instruments to measure student progress towards these standards. The Oregon Statewide Assessment System allows teachers, parents, students and others to compare student performances within a school, within a district, across the state and over time (Oregon Department of Education, 2001).

The Oregon Educational Act for the 21st Century lead to the creation of a three-part assessment system. The three tools of this assessment system are: knowledge and skills tests, performance assessments, and classroom work samples. At least one of the three tools is used in each content area (reading, writing, speaking, mathematics, science, social sciences, arts, second languages) each year to measure student achievement. In 1996 a panel of national experts reviewed Oregon's content and performance standards, and reported that the standards were rigorous. In 1997, *Education Week*, a national educational newspaper, awarded Oregon an A- for its standards and assessment system in a review of public education in all 50 states.

The Oregon state tests include about 70 knowledge and skills reading/literature questions and about 50 knowledge and skills questions in mathematics, science and social sciences. A certain percentage of the questions come from each category in which the scores are reported (see Table 4.7). Scaled scores for the statewide student achievement data are computed using the Rasch model for measurement, based on the work of Benjamin Wright and his students at the University of Chicago (Wright, 1985). Rasch measurement was introduced to the Oregon Department of Education by George Ingebo, a former research director for Portland Public schools. Ingebo demonstrated the advantages of using Rasch methods over conventional statistics for assessing student response to basic skills testing (Ingebo, 1997). The Oregon scaling methods were further refined by Dr Fred Foster, and were adopted when Oregon created its statewide assessment system. Anchor points were established, and the scaling has been periodically refreshed (E. Bigler, personal communication 2001).

Student answer sheets are scanned by computer at a testing company hired by the Oregon Department of Education. Comparing the scanned results with the answer key gives a raw score. The testing company then uses a table to convert the raw score into a scaled score, called a Rasch Unit or RIT. Students receive scaled scores ranging from 150 RITs to about 300 RITs. The RIT scale is based on the convention that 10 RITs equal one logit (Rasch log odds unit). Benchmarks have been set for student results that meet and exceed the state standard for each subject and grade level (see Table 4.8).

Table 4.7
Oregon Statewide Test score reporting categories

Reading & Literature		Mathematics		Science	
Word meaning	18 %	Calculations and estimations	20 %	Unifying concepts and processes	20 %
Literal comprehension	18 %	Measurement	20 %	Physical science	20 %
Inferential comprehension	18 %	Statistics and probability	20 %	Life science	20 %
Evaluative comprehension	18 %	Algebraic relationships	20 %	Earth science	20 %
Locating information	10 %	Geometry	20 %	Scientific inquiry	20 %
Literary forms	7 %				
Literary elements and devices	11 %				

Table 4.8
OSS benchmarks

	Benchmark (RITs)	
	Meets standard	**Exceeds standard**
Mathematics / Reading & Literature:		
Grade 5	215	231
8	231	239
10	239	249
Science:		
Grade 5	223	239
8	233	247
10	239	252

Data were made available by the Oregon Department of Education for the STEP students' achievement in Mathematics, Reading & Literature and Science throughout the period of this study. Additional data sets (May 1999) for 5th-7th grade students provided information about achievement levels prior to the start of the STEP intervention. Mathematics and Reading & Literature are tested annually in Oregon. However, the

Science is tested only at 5th, 8th, and 10th grades. It is unfortunate for the purpose of this study that annual statewide testing in science does not take place at all grade levels in Oregon. Change over time in science could be followed for the 8th grade STEP cohort only (tested in 2000 and 2002), as a result of the limited availability of science test data. Using other in-school achievement results to investigate science achievement was not an alternative option to the use of state wide scores, because of the high variability of science course offerings and grading schemes among classes. Teachers in the USA have much more autonomy and freedom of choice as to what is taught and how it is assessed than do their peers in the UK or Australia. It is not unusual for teachers of the same subject and grade within a school to use quite different teaching materials and assessment items.

Data management

The cognitive and achievement data for the Molalla River School District were provided to me each year as Excel files by the Clackamas County Test Scoring & Reporting Services. Over the period of the study, 17 data files of 6th to 11th grade BLOT results and 90 data files of statewide achievement scores were amassed, a total of approximately 15,000 student files. The various types of data were not directly compatible, and building the master data files for this research project was a time-consuming process. Merging old and new data files was rarely an easy option. Careful checking was needed to identify individuals, often requiring detective skills to track down students who had misidentified themselves by giving incomplete identity information on the Scantron answer sheets.

Analyses of the BLOT master file allowed the growth of individual students to be followed over time. Table 4.9 shows the information displayed in the master file for each student's profile. Once data were collated and merged, information that could identify individual students was removed from the data files to ensure anonymity.

Table 4.9
Student profile

Student identity number
Name of student
Grade in September 1999
Sex
Birth date
Ethnicity
Level of parent education
Number of CASE activities experienced
BLOT raw score, Rasch ability estimate and standard error for 99/00/01/02
Mathematics RIT score for 99/00/01/02
Reading and Literature RIT score for 99/00/01/02
Science RIT score (8th grade 99/00/01/02, 10th grade 02 only)

Controls

If the STEP had been a fully funded research project, then longitudinal data for a relevant control cohort would have provided the best comparisons with the various STEP samples. Clearly such data do not exist, because we were not in a position to collect them for this study. By making the best use of the data that were available, pre-intervention performance profiles and student samples from two non-project school districts served as controls in the analysis. The pre-intervention profiles for cognitive development were constructed using the results of the initial 1999 BLOT testing of 6th to 11th grade Molalla River students. Corresponding pre-intervention profiles for Mathematics and Reading & Literature were constructed from the results of state-mandated tests. These cross-sectional control profiles represent base-lines before the Molalla School District engaged in the STEP intervention.

There are limitations to the use of cross-sectional controls, such as those employed in the Molalla River study. For example, if there were other standard-enhancing initiatives taking place simultaneously with the STEP in the Molalla River schools, then it would be very difficult to tease apart the individual effects of each project. According to Dr. Pellens, there were no other such initiatives in the Molalla River School District during the period of the STEP intervention. However, there were many other variables that could not be controlled in the Molalla schools, such as the teachers' attitude to the STEP and the number and nature of CASE activities covered, apart from variables associated with other areas of the curriculum.

Conclusions based on comparison with the non-project schools must necessarily be regarded as more tentative than those based on the pre-intervention Molalla data. The Sequoia and Redwood School Districts were selected on the basis of demographic similarity to Molalla. However, there were many unexplored differences between these schools, for example, school ethos, professional development emphasis, and curriculum offerings, that would also be likely to influence student achievement. Within-school controls, such as those used in the original CASE research, are preferable to across-school comparisons. However, the goal of the Administrators of the Molalla School District was for all students to benefit from the STEP, and the idea of setting up concurrent experimental and control classes was not considered.

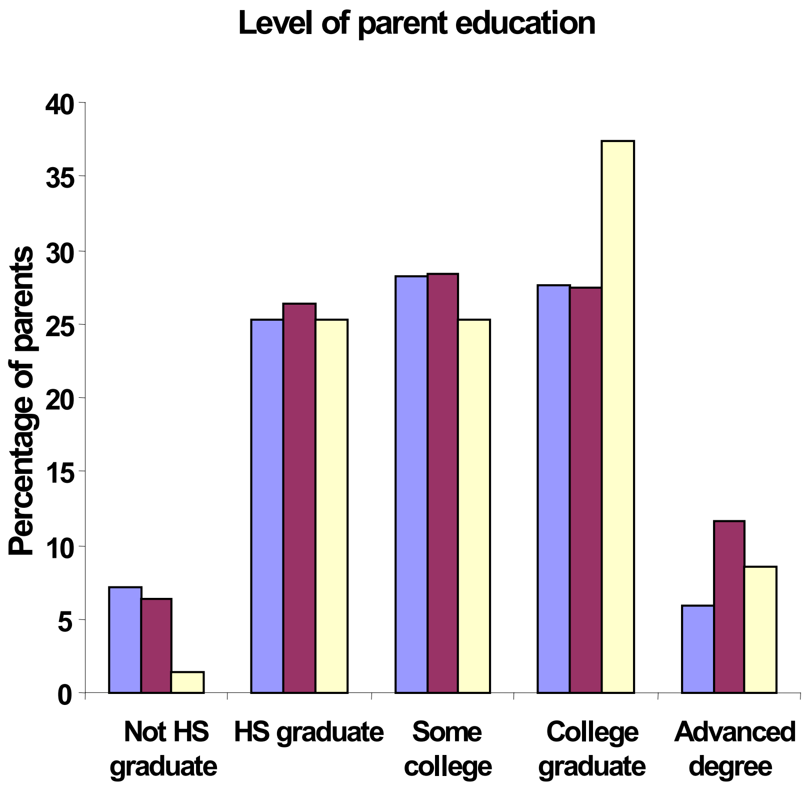
Student achievement data from two other Oregon school districts were provided for comparison with the Molalla River data, and similar master data files were set up for the students of these schools. To ensure anonymity, the comparison school districts have been given the pseudonyms Redwood and Sequoia. These school districts were selected by the administrators of the Molalla River School District because their socioeconomic status was similar to that of Molalla. Secondary analysis of student information data from the state test data files of the comparison schools confirmed their suitability in terms of the level of parent education and ethnicity (see Tables 4.10, 4.11). Redwood and Sequoia students were not exposed to CASE strategies.

Table 4.10
Ethnicity

	Ethnicity of students (%)				
	Native American	**Asian**	**Afro-American**	**Hispanic**	**White**
MolallaRiver	0.00	0.63	0.95	4.74	93.6
Redwood	0.00	0.00	0.33	2.65	97.02
Sequoia	0.68	1.35	0.00	1.35	96.6

Table 4.11
Level of parent education

	Level of parent education (%)				
	Not high school graduate	**High school graduate**	**Some college**	**College graduate**	**Advanced degree**
Molalla River	7.1	25.3	28.2	27.6	5.9
Redwood	6.3	26.3	28.4	27.4	11.6
Sequoia	1.4	25.3	25.3	37.4	8.6



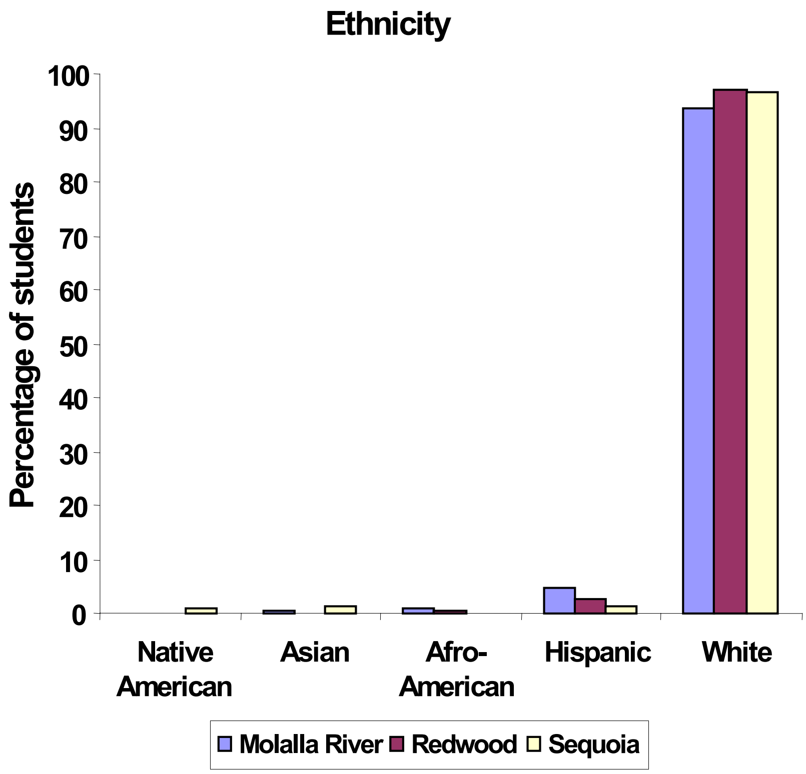


Figure 4.1 Comparison of demographics of Molalla River, Sequoia and Redwood School Districts

Analysis

An innovative feature of this study is that the data from the cognitive test BLOT and the Oregon student achievement tests are both Rasch-scaled. It follows that genuine interval scale measurement is a property of the analysis of the results of the STEP, a characteristic that is not associated with any of the other CASE studies.

The raw scores of the test of cognitive development, BLOT, were subjected to Rasch analysis using the Quest software package (Adams & Khoo, 1993). Persons with extreme scores (all correct or incorrect) are not assigned an ability estimate by the 1993 version of Quest. A beta version of the Quest software (Quest 90 PISA version August 1999), an enhanced version of Quest that uses the Warm estimate, was used to obtain an ability estimate for candidates who scored 35/35 in the 1999/2000/2001/20002 BLOT tests. The extrapolation provided by Quest 90 was 1.14, which is in keeping with Wright's (1998) recommendation that a 'rule of thumb' for extrapolating extreme scores is that no extreme score extrapolation can be less than one logit, and that extrapolations of more than 1.2 logits require convincing justification. Various statistical techniques were then applied to both the measures of cognitive development and student achievement (source: Sokal & Rohlf, 1995). Statistical methods will be described in the context of their use and application in Chapters 5 and 6.

In 1999 the APA Task Force on Statistical Inference published its report in the American Psychologist. The Task Force emphasized that effect sizes (such as Cohen's *d*) should "always" be reported (p. 599), and that "reporting and interpreting effect sizes in the context of previously reported effects is essential to good research" (p. 599). The 5th edition of the APA Publication Manual (2001) states that effect size reporting is "almost always necessary" (p. 25) and that:

> The general principle to be followed . is to provide the reader not only with information about statistical significance but also with enough information to assess the magnitude of the observed effect or relationship
>
> (p. 26).

In light of these current recommendations, effect size (ES) is reported in this study to indicate the magnitude of the effect of the STEP intervention. Effect size measures are now the common currency of meta-analytical studies, and allow comparison of the findings from a specific area of research, such as that involving the various international CASE studies. The effect size measure that is commonly reported in the CASE literature is that of *d*, that is:

$$d = (m_A - m_B) / \sigma$$

where d = ES index for t tests of means,

$m_{A,} m_B$ = sample means, and

o = the average standard deviation of the two samples

In the analysis of the STEP results, the effect size index *d* was estimated using the on-line calculator at www.uccs.edu/~lbecker/psy590/escalc3.htm (Becker, 2003). Cohen (1988) has ascribed threshold values to *d*, that can be seen in Table 4.12.

Table 4.12
Threshold values for effect size index d (from Cohen, 1988, page 82)

THIS TABLE HAS BEEN REMOVED DUE TO
COPYRIGHT RESTRICTIONS

Qualitative methods

Qualitative methods were not used extensively in this study, which was intended to be primarily quantitative and longitudinal in nature. The geographically distant location of the study site from my home made frequent visits impractical. I traveled to the Molalla River School District in 2000 and 2001 to attend seminar presentations and set up the data collection. A third visit at the conclusion of the study in 2003 was proposed, but did not eventuate, because the Oregon Department of Education was in crisis. Indeed, that school year was terminated several weeks early because of financial constraints. By the time that school resumed in the fall, I had returned to Australia to complete this dissertation. In an ideal situation, interviews with the STEP teachers and students would have added a further dimension to this study.

Vignettes of individual students are presented in Chapter 7, in order to illustrate some of the more interesting trends in the data, with pseudonyms have used to preserve anonymity. A survey of teacher satisfaction with the STEP intervention was conducted in the Spring of 2003 as part of the evaluation process of the STEP. Dr. Sandra Pellens, the Director of Instruction of the Molalla River School District, produced the survey and was responsible for its distribution and completion by the STEP teachers. Dr. Trevor Bond and I suggested topics to be included in this questionnaire, that is included as Appendix 2.

CHAPTER 5

The cognitive development of STEP students

Rasch analysis of the 1999 BLOT data

The 1999 BLOT data were subjected to Rasch analysis using Quest software (Adams and Khoo, 1993), and a variety of techniques based on Rasch modelling were used to reveal as much as possible about the interrelationships within the data. The results of the 1999 BLOT item analysis are summarized in Table 5.1 and Figure 5.1.

Table 5.1
Item estimate summary statistics (BLOT 1999)

	Grade (N)	6th (52)	7th (181)	8th (148)	9th (175)	10th (90)	11th (12)	All (658)
Summary of item estimates	Mean	0	0	0	0	0	0	0
	SD	0.83	0.82	0.81	0.73	0.84	0.88	0.76
	SD (adjusted)	0.75	0.80	0.78	0.70	0.79	0.14	0.76
	Reliability estimate	0.83	0.95	0.93	0.93	0.88	0.03	0.98
Infit mean square	Mean	1.0	0.99	1.0	0.99	1.0	0.99	0.99
	SD	0.15	0.12	0.10	0.12	0.13	0.41	0.10
Outfit mean square	Mean	1.0	1.02	1.0	0.99	0.98	1.02	1.0
	SD	0.32	0.25	0.19	0.26	0.32	0.79	0.18
Infit *t*	Mean	-0.02	-0.10	0.02	0.02	0.01	0.06	-0.08
	SD	0.91	1.58	1.16	1.40	0.98	0.77	2.26
Outfit *t*	Mean	0.04	0.18	0.01	0.04	0.01	0.17	0.10
	SD	0.87	1.47	0.98	1.28	0.96	0.77	1.87
Items with zero scores		0	0	0	0	0	0	0
Items with perfect scores		0	0	0	0	0	0	0

BLOT 1999: Item Estimates(Thresholds)in Input Order
(N=658 L=35 Probability Level=0.50)

	ITEM NAME	THRSH	ERR	INFT MNSQ	OUTFT MNSQ
1	item 1	-0.80	.11	1.03	1.00
2	item 2	-0.99	.11	1.03	1.13
3	item 3	0.55	.09	1.10	1.11
4	item 4	-0.65	.10	1.01	1.13
5	item 5	-0.24	.10	1.10	1.20
6	item 6	-1.93	.15	0.96	0.84
7	item 7	-0.52	.10	1.02	1.15
8	item 8	0.65	.09	0.92	0.93
9	item 9	0.38	.09	0.96	0.94
10	item 10	-0.09	.10	0.91	0.81
11	item 11	-0.08	.10	0.93	0.88
12	item 12	-1.33	.12	0.89	0.67
13	item 13	0.70	.09	1.14	1.16
14	item 14	-0.72	.11	0.94	1.04
15	item 15	0.86	.09	1.07	1.17
16	item 16	-0.70	.10	1.10	1.10
17	item 17	0.59	.09	0.87	0.80
18	item 18	-0.37	.10	0.95	0.99
19	item 19	0.31	.09	0.94	0.87
20	item 20	-0.28	.10	0.99	1.01
21	item 21	1.38	.09	1.25	1.51
22	item 22	-0.20	.10	0.96	0.96
23	item 23	0.19	.09	1.12	1.26
24	item 24	-0.71	.10	0.94	1.01
25	item 25	0.26	.09	0.98	0.95
26	item 26	0.45	.09	0.94	0.88
27	item 27	-0.78	.11	0.86	0.66
28	item 28	0.94	.09	0.95	0.96
29	item 29	0.29	.09	0.84	0.79
30	item 30	1.05	.09	1.19	1.28
31	item 31	0.57	.09	1.00	1.08
32	item 32	1.39	.09	1.00	1.03
33	item 33	-0.47	.10	0.90	0.85
34	item 34	-0.12	.10	0.96	0.95
35	item 35	0.41	.09	1.09	1.06
Mean				0.99	1.00
SD				0.10	0.18

Figure 5.1 Item estimates (thresholds) in input order

The reliability indices of the test item data are generally very high (on a 0-1 scale), which means that it is highly likely that the order of difficulty of the item estimates would be replicated if BLOT were to be given to other suitable samples of students (Bond & Fox, 2001). The only exception is the 11th grade reliability index, that was derived from the test results of a very small sample of students (N=12).

There were no items with zero scores (too hard) for this sample, and no items with a perfect score (too easy), which shows that all 35 BLOT items were useful and discriminating for this sample. The mean item fit statistics were also satisfactory. The test summary mean infit and outfit *t* values, which range from -0.10 to +0.18, show that the majority of item estimates have fit values well within the conventionally acceptable range of -2.0 to +2.0 (see Table 5.1). The fit statistics associated with the particular item estimates are displayed in Figure 5.1.

An item-person map (Figure 5.2), generated from the 1999 data, showed questions 6 and 12 to be the easiest items, and 21 and 32 to be the most difficult. Figure 5.2 shows the performance of those students in 6th-11th grades who were pre-tested with BLOT and scored 34/35 or less. A beta version of the Quest software (Quest 90 PISA version August 1999), an enhanced version of Quest that uses the Warm estimate, was used to obtain an ability estimate for candidates who achieved a perfect score in BLOT. Students with perfect scores are not shown on the map, but are estimated to be located 1.14 logits above the students with a raw score of 34/35 (*i.e.* at 4.91 logits).

```
BLOT Ability Estimates                 BLOT Item Difficulty Estimates
----------------------------------------------------------------------
  4.0                             |
                                  |
                     XXXXXXXXX    |
                                  |
                                  |
                                  |
                                  |
  3.0          XXXXXXXXXXXXXXX    |
                                  |
                                  |
                    XXXXXXXXXX    |
                                  |
              XXXXXXXXXXXXXXXX    |
                                  |
  2.0            XXXXXXXXXXXXX    |
              XXXXXXXXXXXXXXXX    |
                                  |
     XXXXXXXXXXXXXXXXXXXXXXXXX    |
               XXXXXXXXXXXXXXX    |     21     32
              XXXXXXXXXXXXXXXX    |
             XXXXXXXXXXXXXXXXX    |     30
  1.0         XXXXXXXXXXXXXXXX    |     28
               XXXXXXXXXXXXXXX    |     15
                XXXXXXXXXXXXXX    |      8     13     17
                      XXXXXXXX    |      3     26     31
                   XXXXXXXXXXX    |      9     19     29     35
                   XXXXXXXXXXX    |     23     25
  0.0         XXXXXXXXXXXXXXXX    |
                     XXXXXXXXX    |     10     11     34
                     XXXXXXXXX    |      5     20     22
                       XXXXXXX    |     18
                     XXXXXXXXX    |      7     33
          XXXXXXXXXXXXXXXXXXXX    |      4     14     16     24
                        XXXXXX    |      1     27
 -1.0                             |      2
                         XXXXX    |
                          XXXX    |
                         XXXXX    |     12
                            XX    |
                                  |
                           XXX    |
 -2.0                             |      6
                                  |
                            XX    |
                                  |
                             X    |
                                  |
                                  |
 -3.0                             |
                             X    |
                                  |
                                  |
                                  |
                                  |
                                  |
 -4.0                             |
----------------------------------------------------------------------
```

Figure 5.2 BLOT item-person map for Molalla River students (1999).
Items are indicated by the item number, and each X represents two students. Students and items are located according to their respective person ability and level of item difficulty.

The person distribution spans almost seven logits (-3.06 to 4.84), which indicates the wide ability range across the 6th to 11th grade student population. In contrast, the item distribution spans only 3.3 logits (-1.8 to 1.5). This means that the locations of 215 students above 1.5 logits are discriminated by only the two most challenging questions, 21 and 32. If the BLOT were to be used primarily to discriminate among students in the upper grades of secondary school, the efficacy of the test would need to be improved by the addition of some questions harder than 21 and 32. Fortunately, most of the STEP participants were relatively younger, and less developed, students from the Molalla Middle School and lower grades only of Molalla High School, and so were reasonably well targeted by the BLOT test.

The person fit statistics generated by the Rasch analysis of the entire 1999 data set show a reliability index of 0.86, and infit and out fit *t* values (0.72, 0.67) well within the acceptable range of -2 to +2. Only 4 out of 658 students had under-fitting statistics marginally outside of this range, indicative of erratic performances (see Table 5.2).

Table 5.2
Ability estimate statistics of four under-fitting students (BLOT 1999)

Student	Grade	Infit mean square	Outfit mean square	Infit *t*	Outfit *t*
A	7	1.37	2.13*	1.54	2.5*
B	8	1.25	2.52*	0.91	2.55*
C	9	1.48*	1.75*	2.36*	2.13*
D	9	1.28	2.07*	1.37	2.56*

Under-fit identified by *

Student C is an Hispanic English language learner (ELL). It is likely that a language dependent test such as BLOT is not the most appropriate method for testing logical operations with ELL students. There is also the possibility of cultural bias, because children in Mexico have been exposed to a quite different environment and lifestyle from their US counterparts. This Hispanic student, together with the other three under-fitting students (not ELL), scored well on several of the most difficult questions, yet had incorrect answers for some of the easiest items. This is a characteristic response pattern known as 'sleeping' that can indicate guessing (Wright, 1979). The infit statistics for the other three students are not anomalous (see Table 5.2). It is likely that ability level has been underestimated for these three students, whereas the result for the ELL child is inaccurate.

In order to ensure the uniformity of the measuring instrument over time, the results of all the subsequent BLOT test analyses were anchored to the 1999 item values. This caused the level of difficulty of the items in the 2000, 2001 and 2002 tests to be anchored to the values generated by the pre-test, and ensured that the same "ruler" was used to measure cognitive development on each occasion. However, in practical terms, anchoring caused very little change in BLOT person ability estimations: the majority did not change, and the others by no more than 0.01 logits.

Rasch analysis of the 2000, 2001 and 2002 BLOT data

Three additional annual rounds of BLOT testing took place in May of 2000, 2001 and 2002. Rasch analysis of these subsequent BLOT tests also showed satisfactory patterns of reliability and good fit. A comparison of the results of the four BLOT tests are displayed in Table 5.3. Test-retest comparisons of scores on the BLOT could be held to be indicators of genuine cognitive development only to the extent that the BLOT test itself is free from what are commonly known as retest effects. Bond's original (1976) report of the BLOT provides convincing evidence of the lack of any retest effect.

Table 5.3
Ability estimate statistics (BLOT 1999, 2000, 2001, 2002)

	09/1999	05/2000	05/2001	05/2002
Grades tested	**6th - 11th**	**6th - 9th**	**7th - 9th**	**7th - 10th**
Number of students	658	446	437	454
Mean of ability estimates[1]	0.91	0.76	0.89	1.16
SD ability estimates	1.22	1.17	1.17	1.25
Reliability	0.85	0.87	0.86	0.86
Infit *t* SD	0.73	0.83	0.71	0.66
Outfit *t* SD	0.67	0.70	0.64	0.64
Zero scores	0	0	0	0
Perfect scores	7	2	2	10

[1] Item estimates anchored to 1999 values

Master data file

Student participation in the data collection procedures for BLOT was completely a matter for the classroom teachers and the students who were in the classes when the intervention was implemented and the data collected. The master file of results was based on the total number of students who had completed the STEP intervention and had been tested with BLOT at least once in the 1999-2000 school year, and again on at least one other subsequent occasion.

The vagaries of day-to-day educational procedures including test taking, record keeping, absenteeism, and changes in the school population, meant that the master data file contained some empty data cells. In order to make a thorough examination of the possible effects of the STEP intervention on cognitive development, a number of data analytical perspectives were taken. For every comparison, the maximum complete data set (*i.e.* all those students without relevant missing data cells) was included in the calculation. For the estimation of change over time, and the subsequent comparison with the control data, students included in the growth estimate were the children in the relative cohort who were tested at both the beginning and the end of the appropriate period. Therefore, the sample sizes for the periodic growth comparisons (students tested twice in 12 months) seen, for example, in Table 5.6, are necessarily smaller than those in Table 5.5, that shows the mean results for children tested merely twice in the 32-month period.

The Molalla River control

If the STEP had been a fully funded research project, then longitudinal data for a control cohort would have been available for comparison with the BLOT results of the STEP students. In the absence of such data from a control cohort, the mean ability estimates of students who were tested with BLOT in September 1999, prior to the STEP intervention, were used to create a cross-sectional control profile to provide comparisons with each of the STEP cohorts (Table 5.4, Figure 5.3).

Table 5.4
Mean ability estimates of Molalla River students by grade (BLOT 1999)

Grade (N students)	**Average age (years; months)**	**Mean BLOT ability estimate (logits)**	**Standard deviation**
6th (50)	11; 6	-0.34	1.29
7th (167)	12; 6	0.66	1.18
8th (148)	13; 6	0.92	1.15
9th (175)	14 ; 6	1.27	1.19
10th (91)	15; 6	1.35	1.13
11th (12)	16; 8	1.33	1.30

The cross-sectional control profile indicates that the time of greatest cognitive change (a gain of approximately 1 logit) occurs between average age 11+ and 12+ years, an age at which some children might be crossing the threshold from concrete to formal operational thinking (see Figure 5.3). Little cognitive change is evident after age 14. This pattern of growth shown in the Molalla River control data is similar to that described by Shayer and Wylam (1978) for some 14,000 British students, using different indicators of cognitive development.

STEP students tested on a minimum of two occasions

Using the results of students who were tested with BLOT on a minimum of two occasions between September 1999 and May 2002, a comparison was made between the cognitive profile of the pre-intervention control population and that of each of the three cohorts of STEP students. Table 5.5 contains a summary of the means of these students' BLOT ability estimates at any particular time.

For the control samples (blue line in the graphs), testing took place in September 1999, whereas for the STEP cohorts (red, orange and green lines in the graphs), testing took place in September 1999 and in May of 2000, 2001 and 2002. Figure 5.4a reveals that the cognitive profile of Cohort A approximates that of the control, although the STEP students both started and finished at a higher mean cognitive level. In contrast, Cohort B (Figure 5.4b) had two periods of apparently faster growth than the control, between age 12+/13+ and 14+/15+ years, interspersed by a sharp loss. Cohort C (Figure 5.4c) followed the same pattern as the control for the first two years of the intervention before showing an apparent growth spurt at age 15+/16+ years. By May 2002, all three STEP cohorts were at a higher mean cognitive developmental level than the control sample.

To investigate whether the STEP students had changed their cognitive levels over the course of the study, regression analyses were performed on the pre-intervention (1999) and post-intervention (2002) BLOT ability estimates of the STEP students. The results indicated that significant improvement in cognitive level had taken place in all three cohorts of STEP students over the period of the intervention (t[df], p Cohort A: 4.31 [30], <0.005; Cohort B: 8.83 [87], <0.0001; Cohort C: 5.47 [37], 0.001). The effect sizes (d) of these results were respectively 1.57, 1.89 and 1.45 standard deviations, all large according to Cohen's scale (Cohen 1988).

Table 5.5
Mean ability estimates of control and STEP students (BLOT 1999, 2000, 2001, 2002)

		Mean BLOT ability estimate in logits (N students)			
Grade	**Average age (years; months)**	**Control**	**Cohort A**	**Cohort B**	**Cohort C**
6th	**11; 6**	-0.34 (50)	-0.05 (46)		
6th	**12; 3**		0.50 (82)		
7th	**12; 6**	0.66 (166)		0.78 (161)	
7th	**13; 3**		0.99 (80)	1.26 (171)	
8th	**13; 6**	0.92 (140)			0.95 (111)
8th	**14; 3**		1.43 (46)	0.93 (139)	1.01 (75)
9th	**14; 6**	1.27 (175)			
9th	**15; 3**			1.91 (103)	1.24 (92)
10th	**15; 6**	1.35 (91)			
10th	**16; 3**				1.87 (45)
11th	**16; 8**	1.33 (12)			

The STEP cohorts were tested in September 1999, May 2000, May 2001 and May 2002, and the cross-sectional control sample in September 1999. In order to compare overall cognitive change in the experimental and control groups, it was necessary to interpolate BLOT ability estimates to the appropriate ages for each control cohort

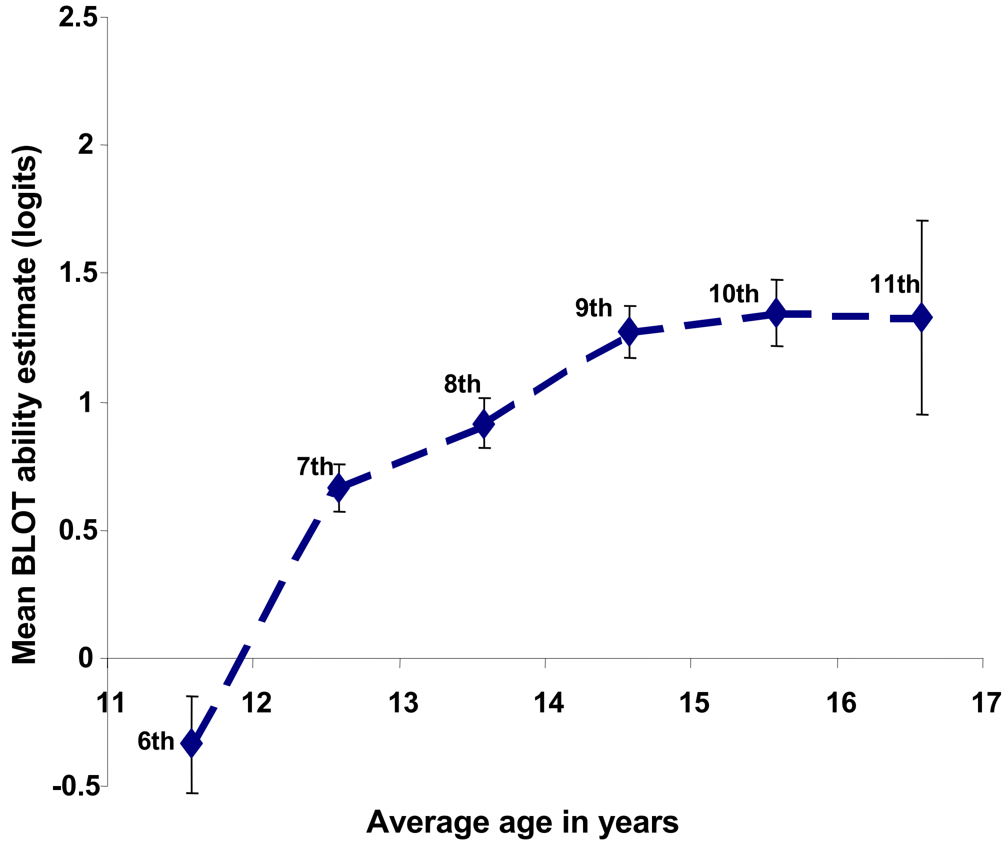


Figure 5.3 Mean BLOT ability estimates of control sample

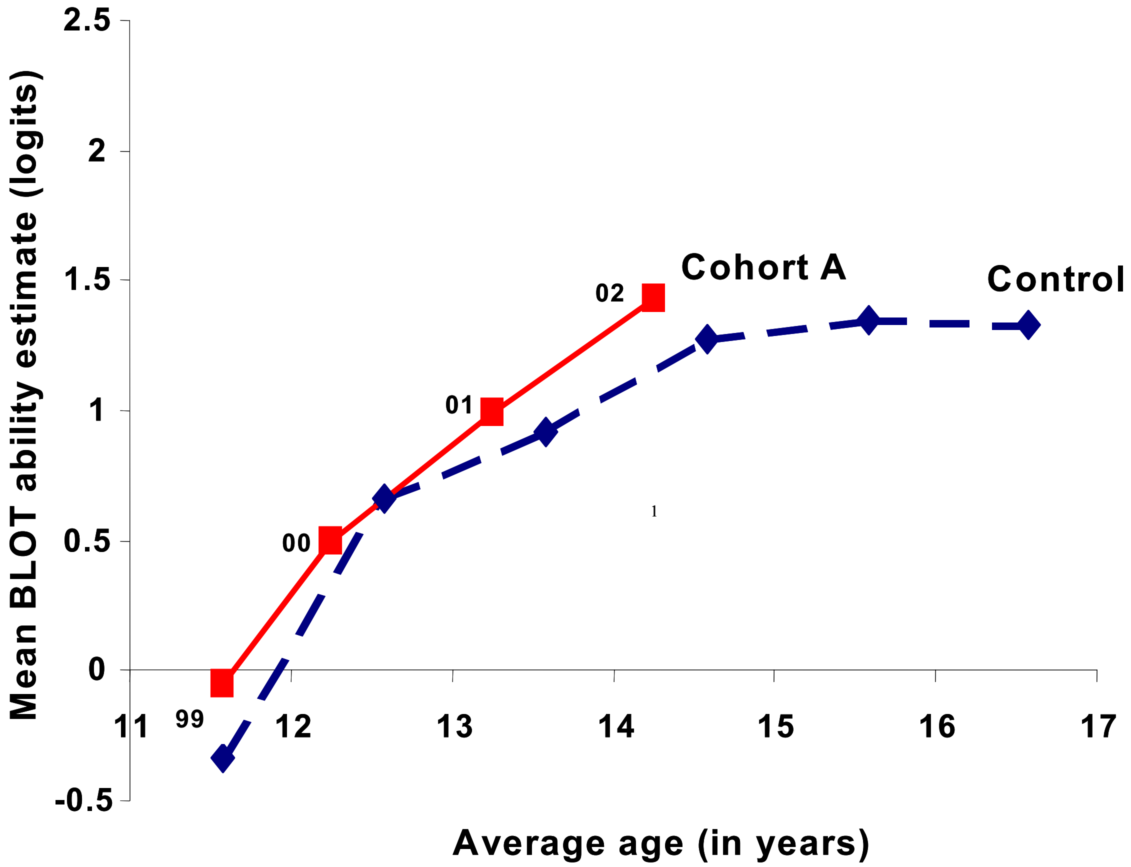


Figure 5.4a Mean BLOT ability estimates of Cohort A

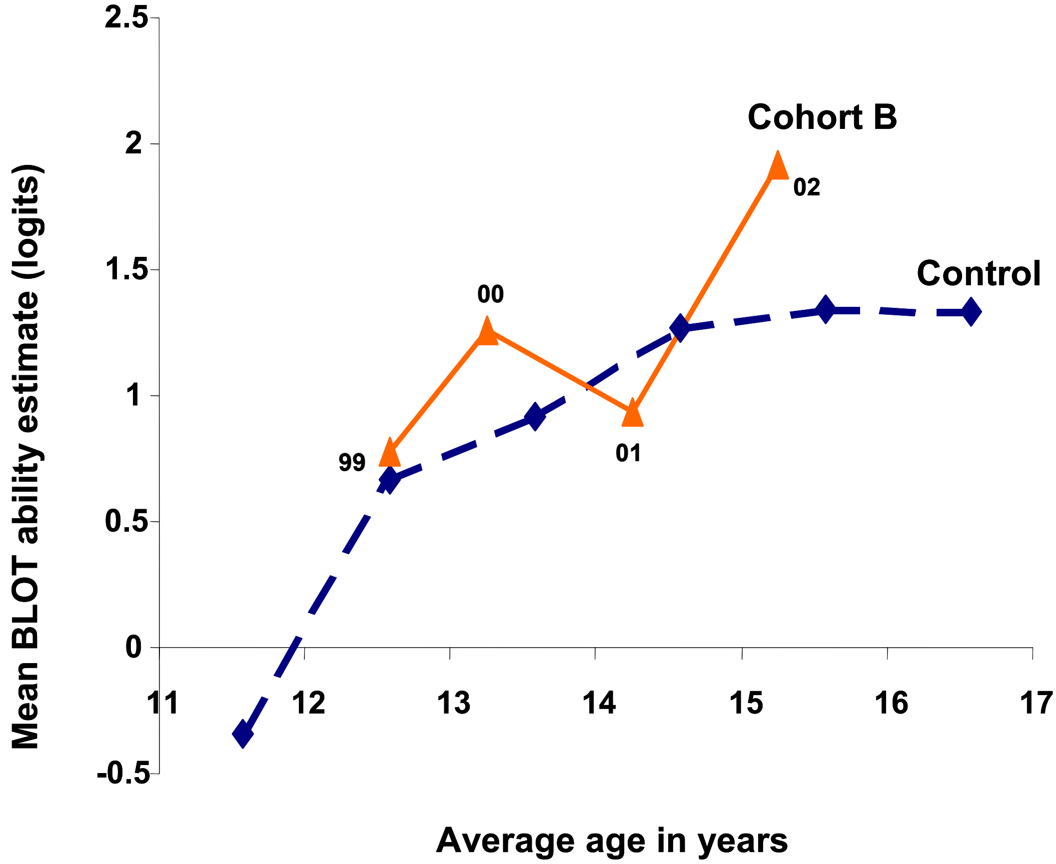


Figure 5.4b Mean BLOT ability estimates of Cohort B

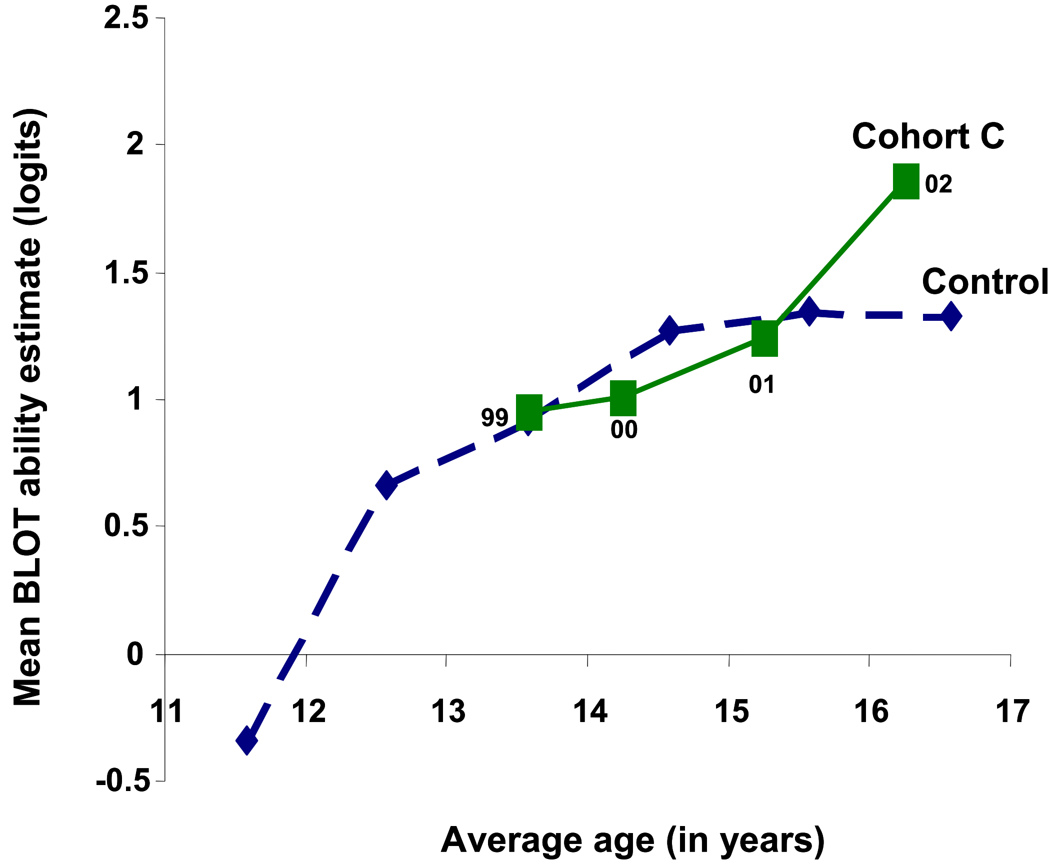


Figure 5.4c Mean BLOT ability estimates of Cohort C

equivalent, because the testing dates were not the same. For example, corresponding mean BLOT ability estimates for the control cohort equivalents in May of each year were interpolated from the data displayed in Figure 5.3, and used in the subsequent analysis. These data were used to calculate the mean change in BLOT ability estimate over each appropriate time period and the overall slope (change/month) for the control sample that are shown in Table 5.6.

Table 5.6
Mean change in BLOT ability estimates of the control sample (from interpolated data)

		Mean change in BLOT ability estimate in logits		
Cohort equivalent	**Overall slope** (change in BLOT ability estimate/time in months)	**0-8 months**	**8-20 months**	**20-32 months**
A	0.0384	0.65	0.45	0.40
B	0.0206	0.16	0.34	0.16
C	0.0134	0.26	0.17	0.07

Table 5.7
Mean change in BLOT ability estimates of STEP cohorts (1999-2002)

	Mean change in BLOT ability estimate in logits			
STEP cohorts	**Overall slope** (change in BLOT ability estimate/time in months	**0-8 months**	**8-20 months**	**20-32 months**
A	0.0366 (76)	0.32 (39)	0.50 (66)	0.50 (34)
B	0.0299 (105)	0.47 (86)	-0.16 (92)	0.83 (96)
C	0.0194 (97)	0.03 (39)	0.36 (46)	0.68 (38)

To investigate overall change for the STEP cohorts, the overall slope (change in BLOT ability estimate over time in months) was calculated for each student and used in the subsequent analysis. Then, the mean and standard deviation of these slopes were calculated for each cohort of STEP students. The mean slope for each cohort equivalent of the control was used as a parameter in *t*-tests for the difference between the overall slopes of experimental and control samples. One-tailed probabilities were used in the *t*-tests because there is an *a priori* expectation that the treatment has a positive effect (Adey & Shayer, 1990, 1994; Iqbal & Shayer, 2000, Endler & Bond, 2001; Kuusela. & Hautamäki, 2002). This procedure was then repeated for the periods 09/1999-05/2000, 05/2000-05/2001, and 05/2001-05/2002 in order to investigate change in each interval between BLOT testing. For these periodic changes, the paramater used was the mean change in BLOT ability estimate, rather than the slope (see Table 5.7).

Of the three STEP cohorts displayed in Table 5.8, only B showed a statistically significant overall cognitive gain (0.56 SDs) against the control from September 1999 - May 2002 [t(df), p = 2.89 (106), <0.01]. However, both Cohorts B and C had statistically significant gains of large effect size over the control in the last year of the intervention (1.18 and 0.93 SDs, respectively). It is interesting to note that Cohort B made a statistically significant gain (0.73 SDs) over the control during the first year of the intervention, followed by a significant loss (- 0.91 SDs) the following year. This significant downward trend in 2000-2001was later found to be associated more with the performance of males than females in Cohort B. A similar downward trend was shown for males of the same age (13+ to 14+ years) in Cohort C, when results for males and females were analyzed separately (see Figure 5.13c). This suggests that under-performance in the BLOT might be associated with some aspect of male adolescent development at this age.

Table 5.8
Comparison of mean change in BLOT ability estimate between control and STEP students

			Mean change in BLOT ability estimate in logits (N students)		
Cohort		**Overall (09/1999-05/2002)**	**09/1999-05/2000**	**05/2000-05/2001**	**05/2001-05/2002**
A	***t*(df)**	-0.25 (75)	-2.40 (38)	0.49 (65)	-0.49 (33)
	p	0.40 NS	0.011 NS	0.31 NS	0.31 NS
	Effect size *d*	-0.06 (nil)	-0.78 (medium)	-0.12 (nil)	0.17 (nil)
B	***t*(df)**	2.89 (104)	3.40 (85)	-4.36 (91)	5.79 (95)
	p	0.0023*	0.00051*	< 0.0001*	<0.0001*
	Effect size *d*	0.56, medium	0.73, medium	-0.91, large	1.19, large
C	***t*(df)**	1.08 (96)	-1.16 (38)	1.41 (45)	2.83 (37)
	p	0.14 NS	0.13 NS	0.083 NS	<0.0037*
	Effect size *d*	0.22 (small)	-0.38 (small)	0.42 (small)	0.93, large

The *t*-tests are one-sample tests comparing control and experimental slopes (column1) and annual change (columns 2, 3, 4). ***t*** represents the *t* statistic, **df** represents degrees of freedom, and ***p*** represents probability of the null hypothesis. * indicates significance with the sequential Bonferroni test (∝ = 0.05).

A regression analysis revealed no significant relationship between of the level of parent education of the STEP participants and the change in BLOT ability estimate from 1999 to 2002 (*t*[df], *p* = -0.04 [114], NS). The effect size was negligible (-0.01 SDs). These results suggest that the level of parent education is not associated with cognitive change in this sample of students (see Figure 5.5).

No significant relationship could be found between the STEP students' starting ability in 1999 and change in BLOT ability estimate from 1999 to 2002 (*t*[df], *p* = -1.20 [144], NS). This result suggests that high gains came from the whole range of starting ability of these students (see Figure 5.6). Similarly, no significant relationship was found between the age at which the STEP intervention commenced and the change in BLOT ability estimate from 1999 to 2002 (*t*[df], *p* = -1.50 [144], NS). Again, this result suggests

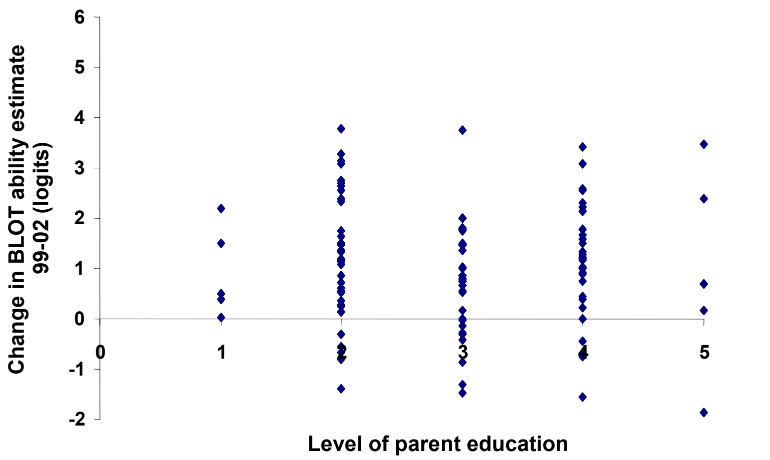


Figure 5.5 Level of parent education and change in BLOT ability estimate

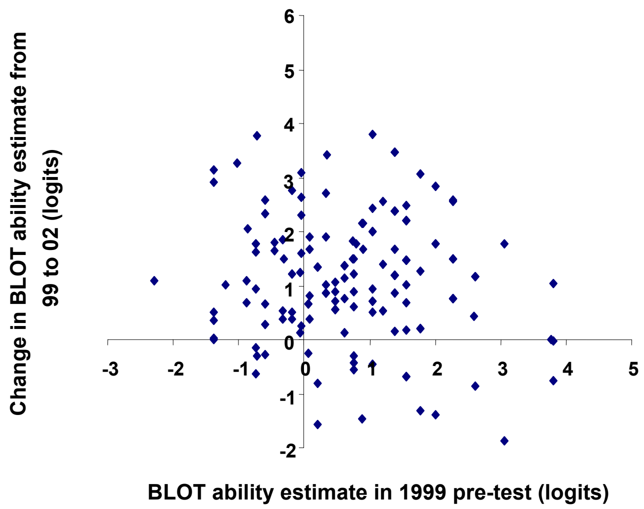


Figure 5.6 Starting developmental level and change in BLOT ability estimate

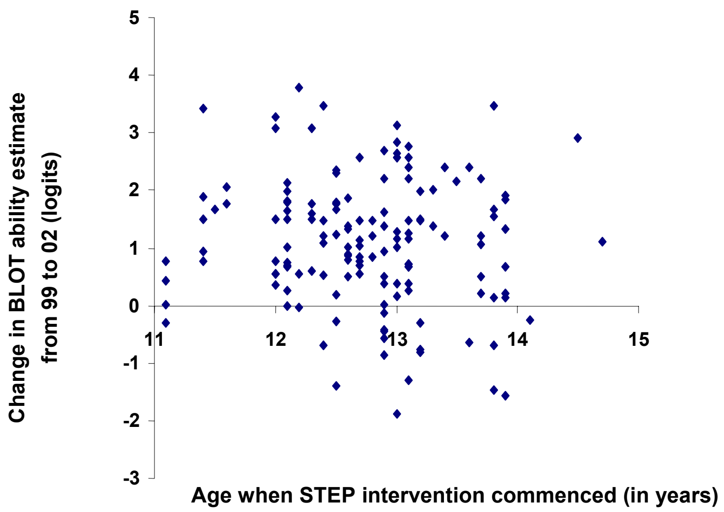


Figure 5.7 Starting age and change in BLOT ability estimate

that high gains were not confined to a specific subset of the student population (see Figure 5.7).

The relative mean of the annual change in the BLOT ability estimate, that takes into account the students' starting level, was also calculated for each cohort of STEP students (see Table 5.9). A positive result for the relative mean indicates that a higher starting value is associated with a greater improvement in that cohort. Whereas a negative relative mean suggests that students who started at a lower ability level improved more than those who started higher. Values around zero indicate no bias towards the starting point.

Table 5.9
Relative means of the annual change in BLOT ability estimate

	Relative mean of the annual change in BLOT ability estimate (logits)		
Cohort	**09/1999 -05/2000**	**05/2000 - 05/2001**	**05/2001 - 05/2002**
A	-0.01	-0.41	0.74
B	0.45	-0.25	-0.79
C	0.05	-0.37	0.15

Different patterns are revealed when the relative means of Cohorts A and C are compared with those of B. There is little bias towards starting level in the first year of the intervention for both Cohorts A and C, followed by the less able students improving faster during the second years of the intervention. By the third year it is the students who started at a higher ability level that are showing more improvement in both Cohorts A and C. Whereas, in Cohort B the higher ability students improved faster in the first year of the intervention, followed by their lower ability peers making more improvement in the second and third years. This suggests that the lower ability students in Cohort B might take longer to benefit from the intervention than their more able peers. It also supports the idea that students from the full range of starting ability can benefit from the STEP intervention, though not necessarily in the same way for all.

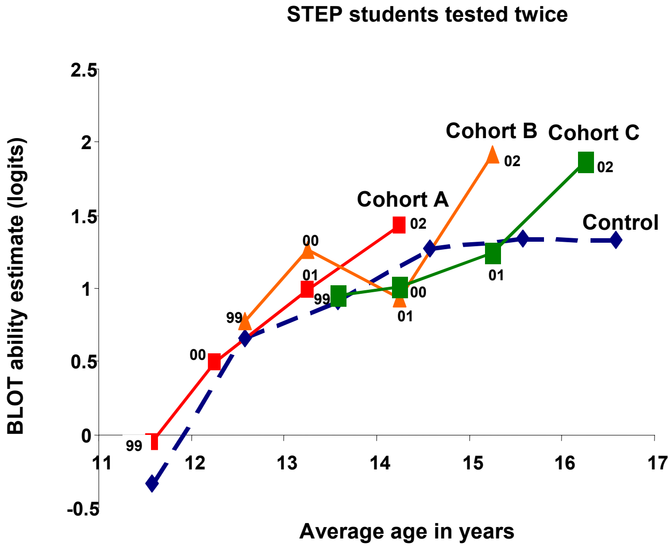
STEP students tested on more than two occasions

The mean BLOT ability estimates of those STEP students tested on three occasions were analyzed as a sub-sample, for comparison with the results for students tested only twice during the program (see Table 5.10).

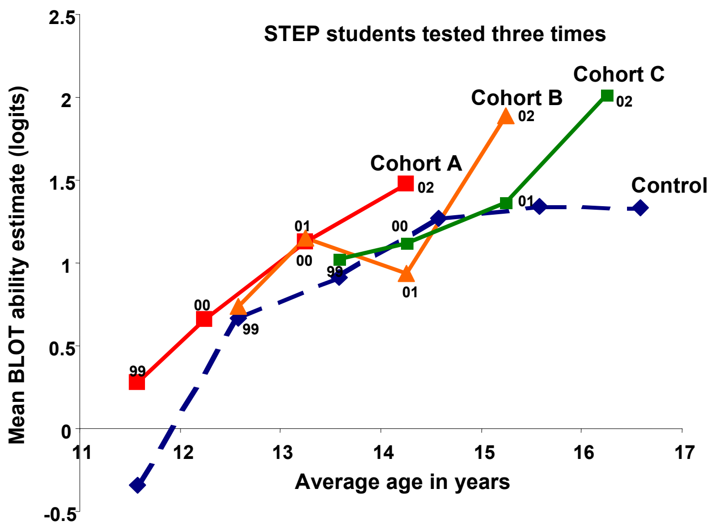
Table 5.10
Mean BLOT ability estimates of students tested on three occasions

Grade	Average age (yrs; mths)	Mean BLOT ability estimate in logits (N students)			
		Control	STEP cohorts		
			A	B	C
6th	**11; 6**	-0.34 (50)	0.27 (36)		
6th	**12; 3**		0.66 (52)		
7th	**12; 6**	0.67 (167)		0.74 (131)	
7th	**13; 3**		1.13 (52)	1.15 (141)	
8th	**13; 6**	0.92 (148)			1.02 (62)
8th	**14; 3**		1.47 (35)	0.94 (123)	1.12 (42)
9th	**14; 6**	1.27 (175)			
9th	**15; 3**			1.89 (118)	1.37 (63)
10th	**15; 6**	1.35 (91)			
10th	**16; 3**				2.02 (39)
11th	**16; 8**	1.33 (12)			

a)



b)



c)

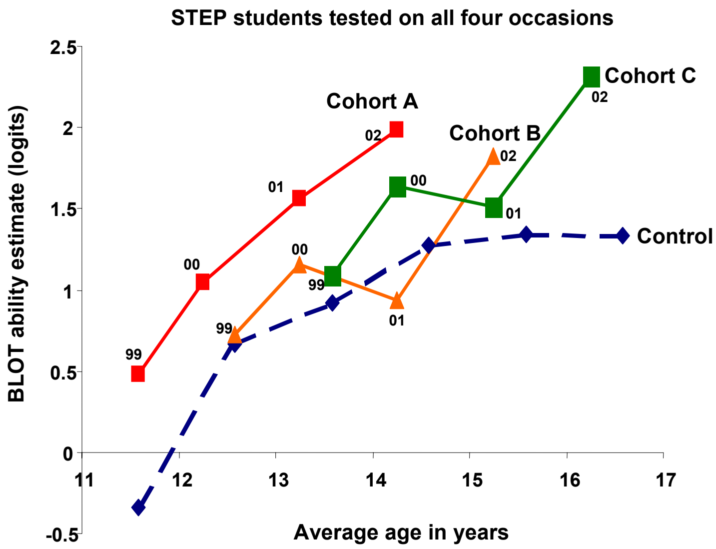


Figure 5.8 a, b, c Cognitive development profiles

Figure 5.8 displays cognitive profiles for STEP students tested on two, three and four occasions. The results for Cohort B shown in Figures 5.8 a and b present a similar pattern, which is not unexpected, given the large overlap between these samples of STEP students. Certain results were slightly elevated for students in Cohorts A and C tested on more than two occasions (*e.g.* the starting point for Cohort A is higher in Figures 5.8 b and c than in 5.8 a), suggesting that children who presented more frequently for testing achieved higher BLOT scores. The results for individual students tested on all four occasions (September 1999, May 2000, May 2001, and May 2002) were analyzed separately to allow for the progress of students to be traced as accurately as possible over the thirty-two months of the program. The mean BLOT ability estimates of the subset of STEP students tested on all four occasions are displayed in Table 5.11 and Figure 5.8c.

Table 5.11
Mean BLOT ability estimates of students tested on all occasions

		Mean BLOT ability estimate (logits)			
Grade	**Average age (yrs; mths)**	**Control**	**Cohort A (N = 16)**	**Cohort B (N = 77)**	**Cohort C (N = 12)**
6th	**11; 6**	-0.34 (50)	0.48		
6th	**12; 3**		1.05		
7th	**12; 6**	0.66 (167)		0.73	
7th	**13; 3**		1.56	1.16	
8th	**13; 6**	0.92 (148)			1.09
8th	**14; 3**		1.98	0.93	1.64
9th	**14; 6**	1.27 (175)			
9th	**15; 3**			1.82	1.51
10th	**15; 6**	1.35 (91)			
10th	**16; 3**				2.32
11th	**16; 8**	1.33 (12)			

Comparison of Figures 5.8 a, b and c suggests that the subset of Cohort A that was tested on all occasions was an elite sub-group of the whole STEP sample, because their mean ability estimates are approximately 0.5 logits higher than those of the control. Their 8th grade teacher commented that most of these students were smarter than average, but that they varied in their attitude to school and science lessons, as well as in their school grades (personal communication, 2003). She rightly felt that their performance in the BLOT was more likely to be a reflection on their skills, rather than on their attitude or behavior. She said that one of the students does twice as much work as is necessary and was sure to be the 2003 valedictorian. This student was only one of two 8th graders to be recommended for Biology (as opposed to the more general Life Science) as a freshman at high school in 2003. It is possible that the elevated BLOT results of the subset of Cohort A that was tested on all occasions might be due to their better than average school attendance. It is also possible that the home background might be more stable for such students who remained in the same school for the entire period of the intervention. However, this could also be said of Cohorts B and C who were tested four times, neither of which showed such a strong elevation.

A comparison of the mean change in the BLOT ability estimate between the control and STEP students tested on all occasions is shown in Table 5.12. The pre-intervention control means, interpolated from Figure 5.3, were again used as parameters in one-tailed *t*-tests, comparing the control and experimental slopes. An overall statistically significant gain (0.55 SDs) was made by Cohort B over the control (t[df], p = 2.40 [74], <0.01), with statistically significant gains for the first (0.67 SDs) and third periods (1.23 SDs), interspersed by a significant loss (-0.92 SDs) in 2000-2001. Cohort C also showed a statistically significant gain (1.65 SDs) against the control overall (t[df], p = 2.74 [11], <0.02). No statistically significant gains were shown by Cohort A, however, there was a significant loss (-1.04 SDs) in 1999-2000.

Table 5.12
Comparison of the mean change in BLOT ability estimate between control and STEP students tested on all occasions

		Mean change in BLOT ability estimate in logits (N students)			
Cohort		**Overall 09/99-05/02**	**09/99-05/00**	**05/00-05/01**	**05/01-05/02**
A	***t*(df)**	0.18 (15)	-2.02 (15)	0.69 (15)	0.04 (15)
	p	0.43 NS	0.031 NS	0.25 NS	0.48 NS
	Effect size *d*	0.09 (nil)	-1.04 (large)	0.35 (small)	0.02 (nil)
B	***t*(df)**	2.40 (74)	2.93 (74)	-4.01 (74)	5.34 (74)
	p	0.0094*	0.0022*	<0.0001*	<0.0001*
	Effect size *d*	0.55, medium	0.67, medium	-0.92, large	1.23, large
C	***t*(df)**	2.74 (11)	0.95 (11)	-0.79 (11)	2.80 (11)
	p	0.0096*	0.18 NS	0.22 NS	0.0086 NS
	Effect size *d*	1.65, large)	0.57 (medium)	-0.48 (small)	1.69 (large)

The t-tests are one-sample tests comparing control and experimental slopes (column1) and annual change (columns 2, 3, 4). ***t*** represents the t statistic, **df** represents degrees of freedom, and ***p*** represents probability of the null hypothesis. * indicates significance with the sequential Bonferroni test ($\propto$ = 0.05).

The sample size is low (N = 16) for Cohort A because not all 6th grade STEP students were pre-tested with BLOT in September 1999. Furthermore, some 8th grade science teachers opted out of the STEP, further reducing the number of students who were tested on all occasions in Cohorts A and C (N = 12) .

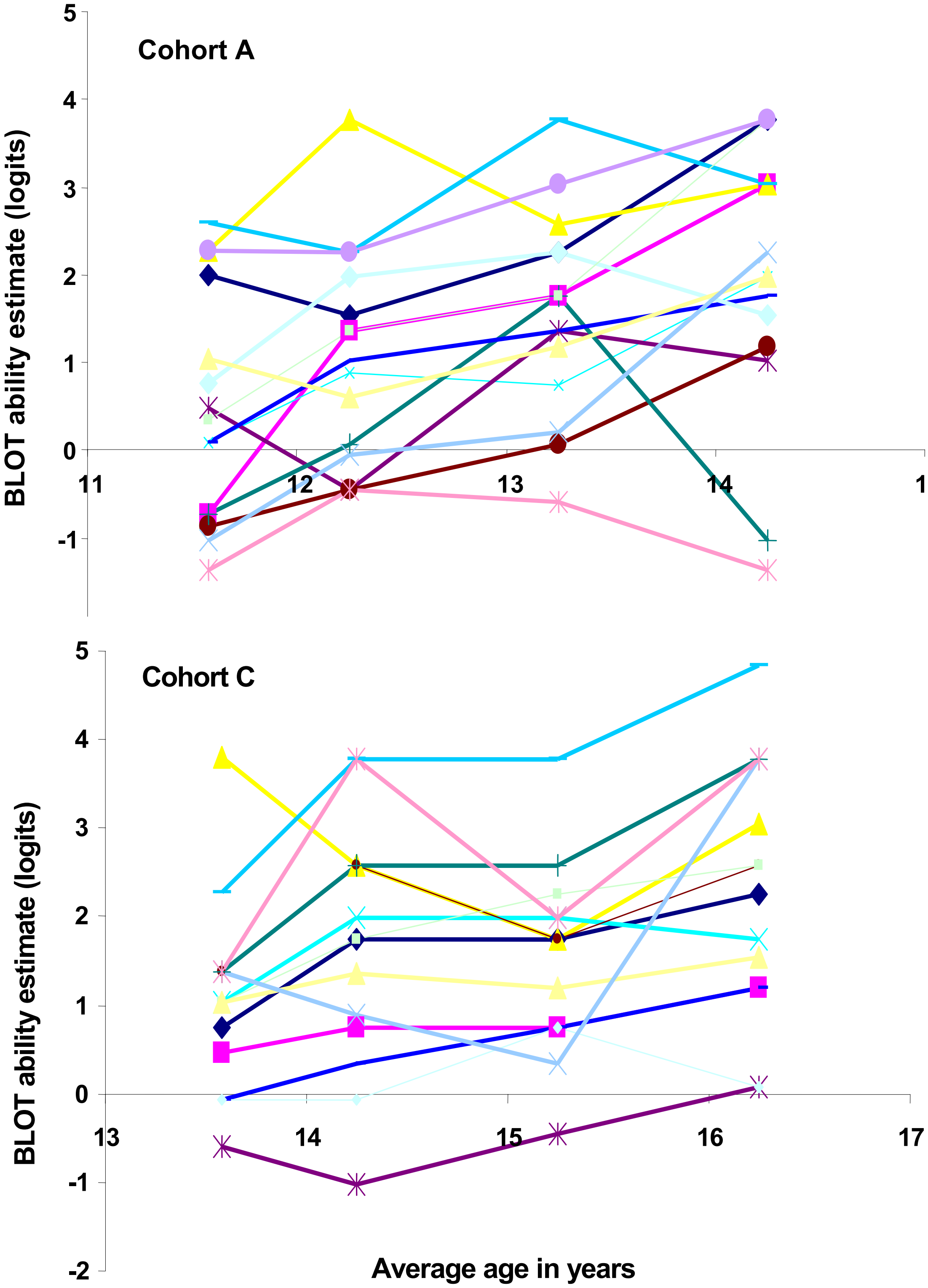


Figure 5.9 BLOT ability estimates of students tested on all occasions

Analysis of the BLOT results for individual STEP students who were tested on all occasions showed that there was remarkable variation in both the amount of cognitive change and when it occurred. Figure 5.9 shows the diversity of individual variation in Cohorts A and C. Similar heterogeneity was found in Cohort B, which was not illustrated due to its much larger sample size. Some students are seen to make little progress when their BLOT ability estimates are examined (*e.g.* 1.21, 1.75, 1.75, 1.75), others are clearly late developers (*e.g.* -0.05, -1.02, -0.59, 3.04), while some show erratic patterns of development (*e.g.* 0.76, 2.25, -0.45, 1.98).

The amount of change in the BLOT ability estimate per year also varied considerably from person to person in all three cohorts of STEP students. This variation is illustrated in Figures 5.10a, b, c, which show histograms of the frequency of the BLOT ability estimates in each of the three years of the program for each cohort of students. There is a general trend of increased numbers of students achieving higher ability estimates over the period of the program, suggesting that students are indeed developing cognitively.

Results for male and female students

The results were analyzed separately for male and female students who were tested at least once in the first year of the STEP intervention and again on a subsequent occasion. The mean BLOT ability estimates for the males and females of each cohort are shown in Table 5.13, together with the means of the pre-intervention control population. Comparison of the male and female students in the control population revealed different patterns of cognitive change for the sexes (see Figure 5.11). The trend for females is that of steady improvement up to age 14+, followed by a steady decline; whereas males

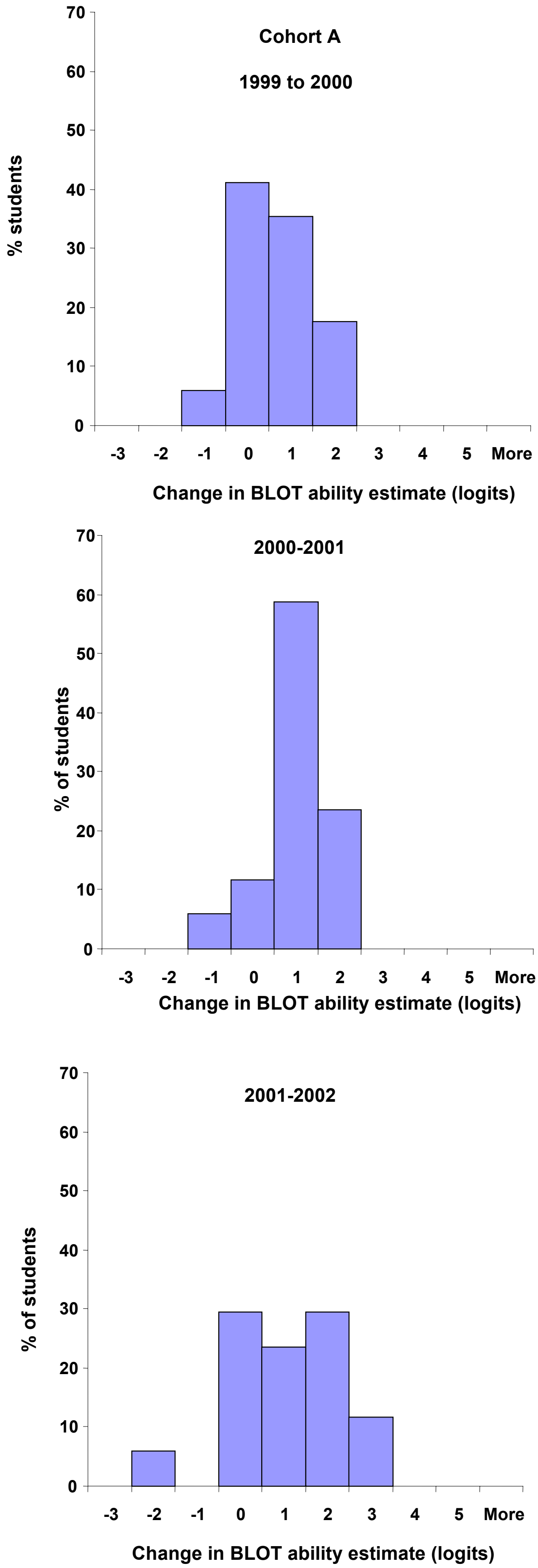


Figure 5.10 a Annual change in BLOT ability estimate of Cohort A

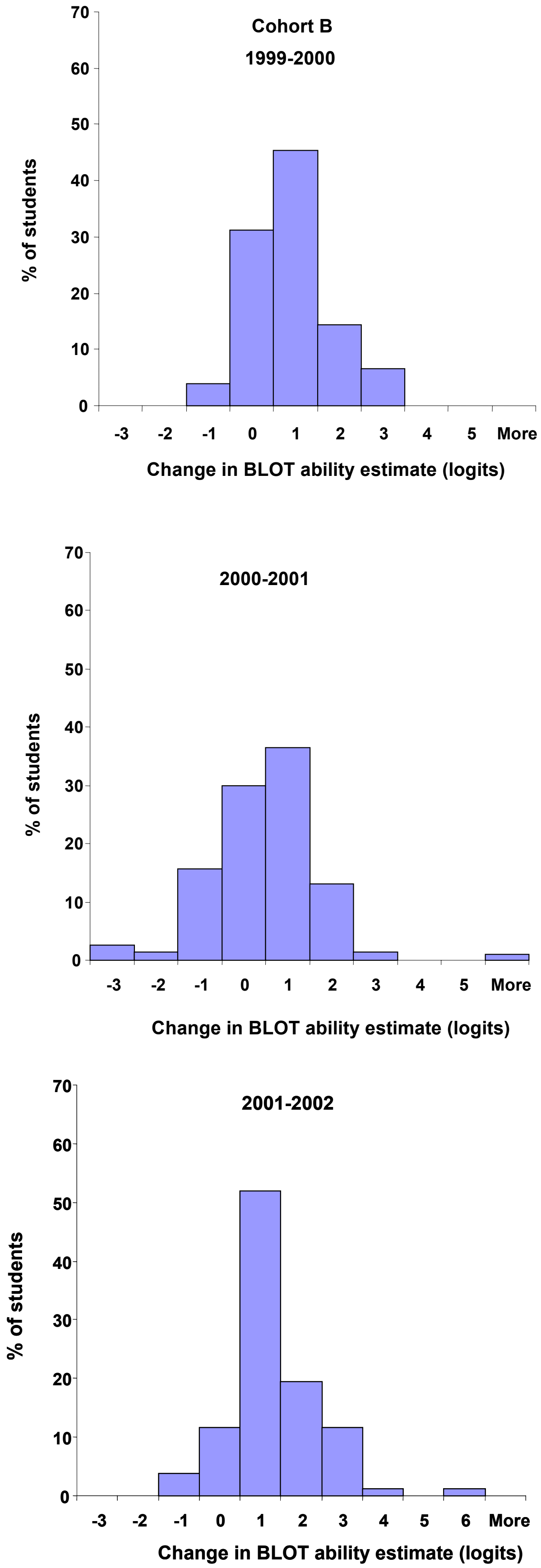


Figure 5.10 b Annual change in BLOT ability estimate of Cohort B

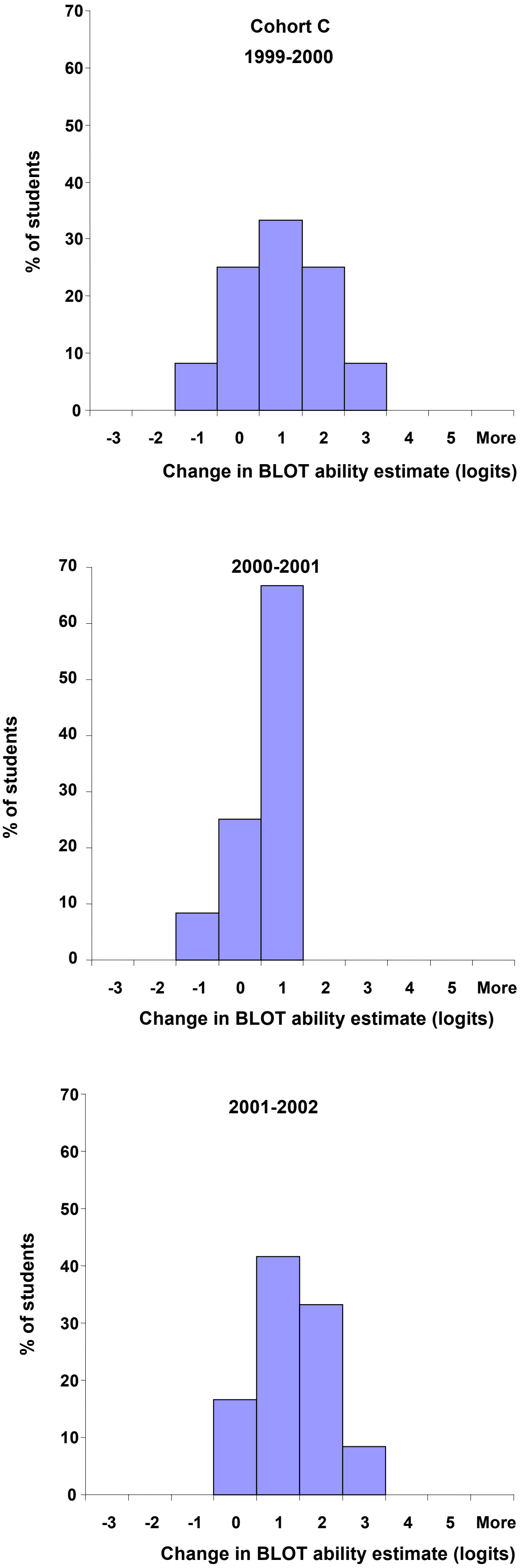


Figure 5.10 c Annual change in BLOT ability estimate of Cohort C

showed a growth spurt at 11+ years, followed by a plateau (12+/13+) and then a second growth phase from age 13+ to16+. The mean ability estimates of male and female control students can be compared with those of the STEP participants in Figure 5.12. The results of one tailed *t*-tests comparing the slopes of female control and STEP students are displayed in Table 5.14.

Table 5.13
Mean ability estimates of male and female students (BLOT 1999, 2000, 2001, 2002)

			Mean BLOT ability estimate in logits (N students)			
Grade	**Average age (years; months)**	**Sex**	**Control**	**Cohort A**	**Cohort B**	**Cohort C**
6th	**11; 6**	male	-0.12 (24)	-0.23 (22)		
		female	0.05 (26)	0.31 (24)		
6th	**12; 3**	male		0.31 (42)		
		female		0.70 (40)		
7th	**12; 6**	male	0.83 (79)		1.17 (78)	
		female	0.51 (87)		0.56 (83)	
7th	**13; 3**	male		0.84 (41)	1.71 (78)	
		female		1.17 (39)	1.04 (93)	
8th	**13; 6**	male	0.84 (71)			0.90 (52)
		female	0.99 (69)			1.00 (59)
8th	**14; 3**	male		1.26 (29)	0.93 (60)	0.57 (36)
		female		1.71 (17)	0.93 (79)	1.37 (39)
9th	**14; 6**	male	1.15 (82)			
		female	1.38 (93)			
9th	**15; 3**	male			2.26 (58)	1.14 (40)
		female			1.74 (75)	1.34 (52)
10th	**15; 6**	male	1.45 (51)			
		female	1.22 (40)			
10th	**16; 3**	male				1.96 (19)
		female				1.78 (26)
11th	**16; 8**	male	1.51 (8)			
		female	0.98 (4)			

One-tailed probabilities were used in the *t*- tests for males and females because there is an *a priori* expectation that the treatment has a positive effect (Adey & Shayer, 1990; Endler & Bond, 2001). No significant difference was found between the slopes of the females in Cohorts A and C over the control (see Table 5.14). In contrast, the slopes for the Cohort B females were significantly different from the control, with a statistically significant gain overall (0.93 SDs). There were also gains for Cohort B in the first (0.78 SDs) and third (2.48 SDs) years of the intervention, interspersed with a loss (-1.43 SDs) in the second year. The corresponding data for male students are shown in Table 5.15. Overall, males made no statistically significant gains over the control. Cohort B males showed a statistically significant gain (1.52 SDs) over the control in 2001-2002, which followed a significant loss (-1.34 SDs) in the previous year.

Table 5.14
Comparison of the mean change in BLOT ability estimate between female control and STEP students

		Mean change in BLOT ability estimates in logits (N students)			
Cohort		**Overall 09/1999-05/2002**	**09/1999-05/2000**	**05/2000-05/2001**	**05/2001-05/2002**
A	***t*(df)**	0.69 (36)	0.47 (20)	-0.18 (31)	0.54 (12)
	p	0.25 NS	0.32 NS	0.43 NS	0.30 NS
	Effect size (*d*)	0.23 (small)	0.21 (small)	-0.06 (nil)	0.31 (small)
B	***t*(df)**	3.83(68)	2.86 (53)	-5.48 (59)	9.68 (61)
	p	0.00014*	0.0030*	<0.0001*	<0.0001*
	Effect size (*d*)	0.93, large	0.78, medium	-1.43, large	2.48, large
C	***t*(df)**	1.04 (50)	-0.65 (21)	1.41 (25)	2.08 (19)
	p	0.15 NS	0.26 NS	0.085 NS	0.026 NS
	Effect size (*d*)	0.29 (small)	-0.28 (small)	0.56 (medium)	0.96 (large)

The *t*-tests are one-sample tests comparing control and experimental slopes (column1) and annual change (columns 2, 3, 4). ***t*** represents the *t* statistic, **df** represents degrees of freedom, and ***p*** represents probability of the null hypothesis. ***** indicates significance with the sequential Bonferroni test ($\propto$ = 0.05).

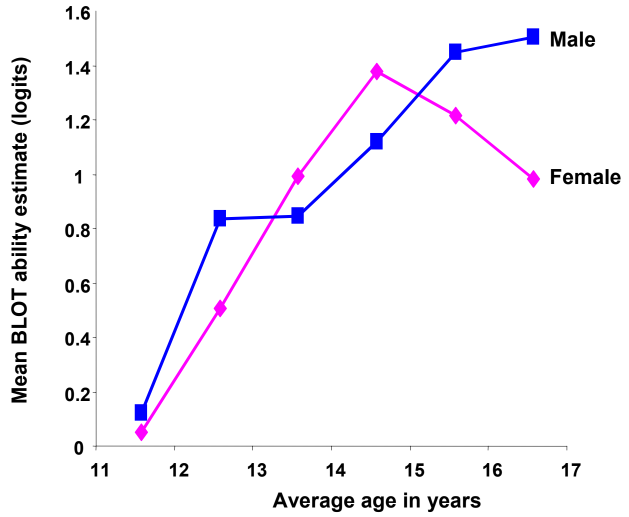
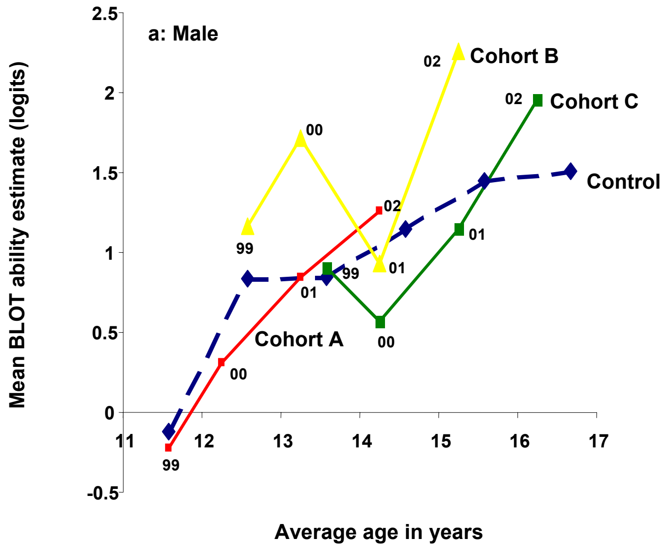


Figure 5.11 Mean BLOT ability estimates of male and female control students



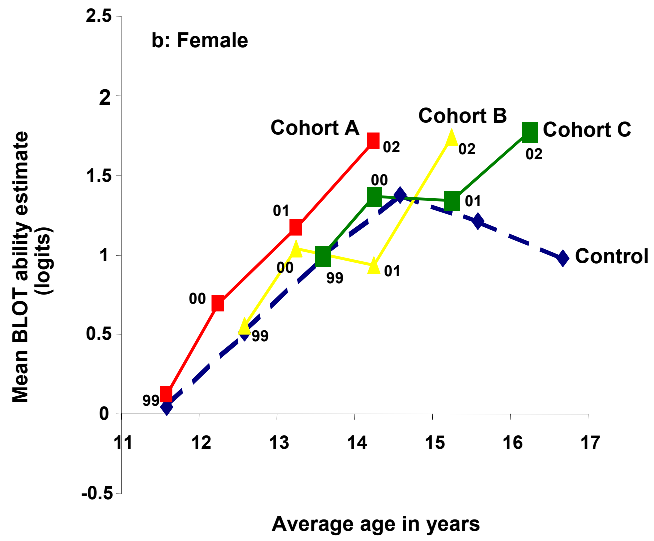


Figure 5.12 Mean BLOT ability estimates of male and female STEP students

Comparisons were also made between the results of males and females in each of the STEP cohorts (see Figure 5.13). No significant difference was found between the BLOT ability estimates of the males and females in any cohort. Female students in Cohort A had a starting point 0.54 logits higher than their males peers, that they sustained throughout the period of the study. The pattern was different for Cohort B, where female students started 0.6 logits and finished 0.53 logits lower than their male peers. The trough that was seen for Cohort B in the 2001 BLOT (see Figure 5.4 b) is now shown in Figure 5.13 to be largely due to the results for the male students, whose mean ability estimate fell by 0.78 logits between 2000 and 2001.

Table 5.15
Comparison of the mean change in BLOT ability estimate between male control and STEP students

		Mean change in BLOT ability estimates in logits (N students)			
Cohort		**Overall 09/99-05/02**	**09/1999-05/2000**	**05/2000-05/2001**	**05/2001-05/2002**
A	***t*(df)**	-0.82 (40)	-2.24 (17)	1.46 (33)	0.35 (20)
	p	0.21NS	0.019 NS	0.077 NS	0.37 NS
	Effect size *d*	-0.26 (small)	-1.08 (large)	0.51 (medium)	0.15 (nil)
B	***t*(df)**	0.75 (38)	0.48 (32)	-3.80 (32)	4.36 (33)
	p	0.23 NS	0.32 NS	0.00031*	<0.0001*
	Effect size *d*	0.24 (small)	0.17 (nil)	-1.34, large	1.52, large
C	***t*(df)**	-0.30 (45)	-0.78 (14)	1.00 (17)	1.57 (16)
	p	0.38 NS	0.22 NS	0.17 NS	0.068 NS
	Effect size *d*	-0.09 (nil)	-0.42 (small)	0.48 (small)	0.79 (medium)

The *t*-tests are one-sample tests comparing control and experimental slopes (column1) and annual change (columns 2, 3, 4). ***t*** represents the *t* statistic, **df** represents degrees of freedom, and ***p*** represents probability of the null hypothesis. * indicates significance with the sequential Bonferroni test (∝ = 0.05).

a)

b)

c)

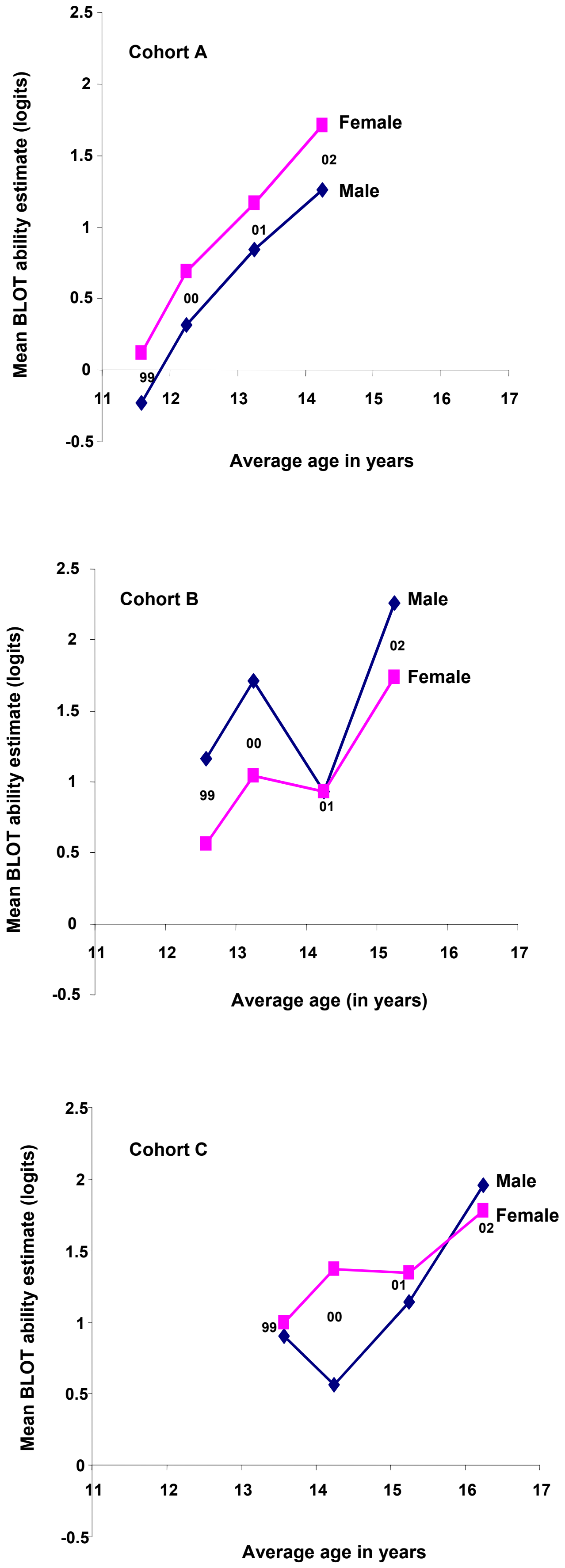


Figure 5.13 Mean BLOT ability estimates of male and female students

In Cohort C, the means for the sexes were close at the start, females only 0.1 logits lower than males, but by the end of the study males had overtaken females by 0.18 logits. From 1999 to 2000, males in Cohort C showed a similar trend to the Cohort B males of the same age (13+ to 14+ years), although the loss was not quite as great at 0.36 logits. This drop in mean BLOT ability estimate for 13/14 year old males in both the Cohorts B and C might indicate that the performance of these males was negatively affected by some aspect of their adolescence or school experience.

The relative mean of the annual change in BLOT ability estimate, that takes into account the students' starting level, was also calculated for each cohort of male and female STEP students (see Table 5.16). Different patterns are revealed when the relative means of males and females are compared. For Cohort A, it is the higher ability males but lower ability females who show most improvement in the first year of the intervention. In the second year, lower ability students of both sexes show improvement. Whereas in the last year it is their higher ability peers who make more gains, with females showing this effect more strongly than males. A similar pattern is shown for Cohort C in the first and second years. However, in the third year, the sexes differ, with more improvement shown by higher ability females and lower ability males. In the last two years of testing, Cohort C females show more improvement than their male peers. There is most similarity between the sexes in Cohort B, where higher ability students of both sexes improve more in the first year, followed by gains for their lower ability peers in the second and third years.

Table 5.16
Relative means of annual change in BLOT ability estimate of male and female STEP students

		Relative mean of annual change in BLOT ability estimate (logits)		
Cohort		**09/1999 -05/2000**	**05/2000 - 05/2001**	**05/2001 - 05/2002**
A	male	0.79	-0.39	0.54
	female	-0.69	-0.44	1.06
B	male	0.11	-0.24	-1.21
	female	0.63	-0.24	-0.56
C	male	0.33	-0.01	-0.24
	female	-0.13	-0.66	0.50

Results for students who experienced only part of the STEP intervention

A subset of students, who later became peers of Cohort A, did not experience the STEP intervention in 1999-2000. Some of these students, who began the STEP in 7th grade (September 2000), were new to the school, whereas others had been in classes that did not participate in the intervention in 6th grade. Table 5.17 and Figure 5.14 show the mean ability estimates of these students, their STEP peers, and the pre-intervention control. Students who experienced the first year of the STEP intervention in 1999-2000 were exposed to CASE activities 1 through 5. The STEP teachers considered these activities, that included controlling variables and the fair test, to be essential to their goal of improving their students' competency in scientific inquiry, and for this reason they revisited these initial CASE activities at the start of each school year. It follows then that the students who did not experience the first year of STEP in 1999-2000, did not actually miss the content of any of the CASE activities, although the material might have been presented in a different manner and delivered over a shorter period of time.

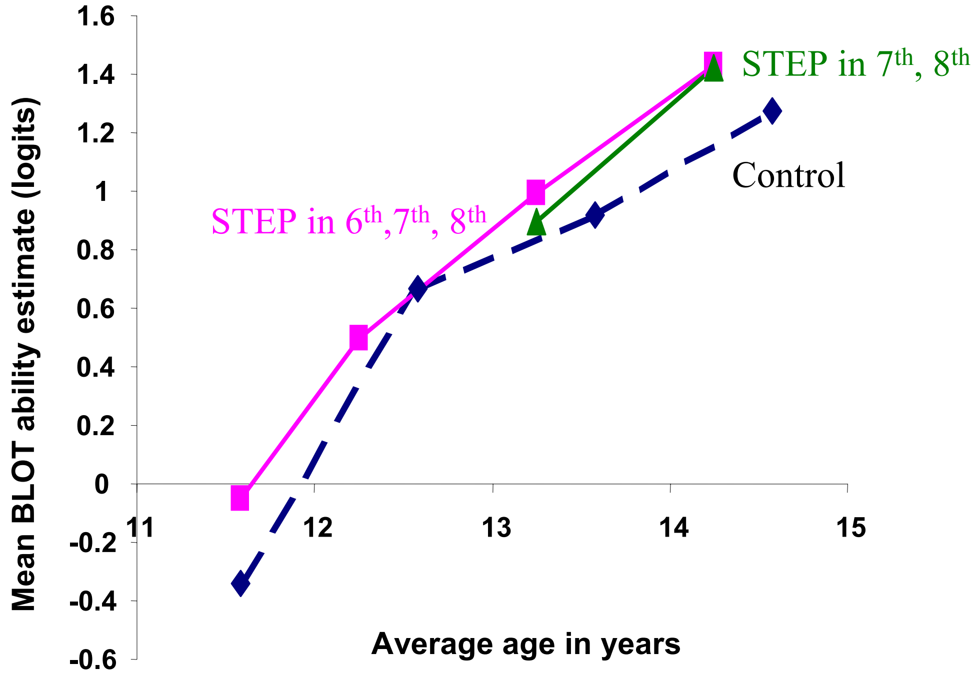


Figure 5.14 Peers of Cohort A who missed STEP in 6th grade

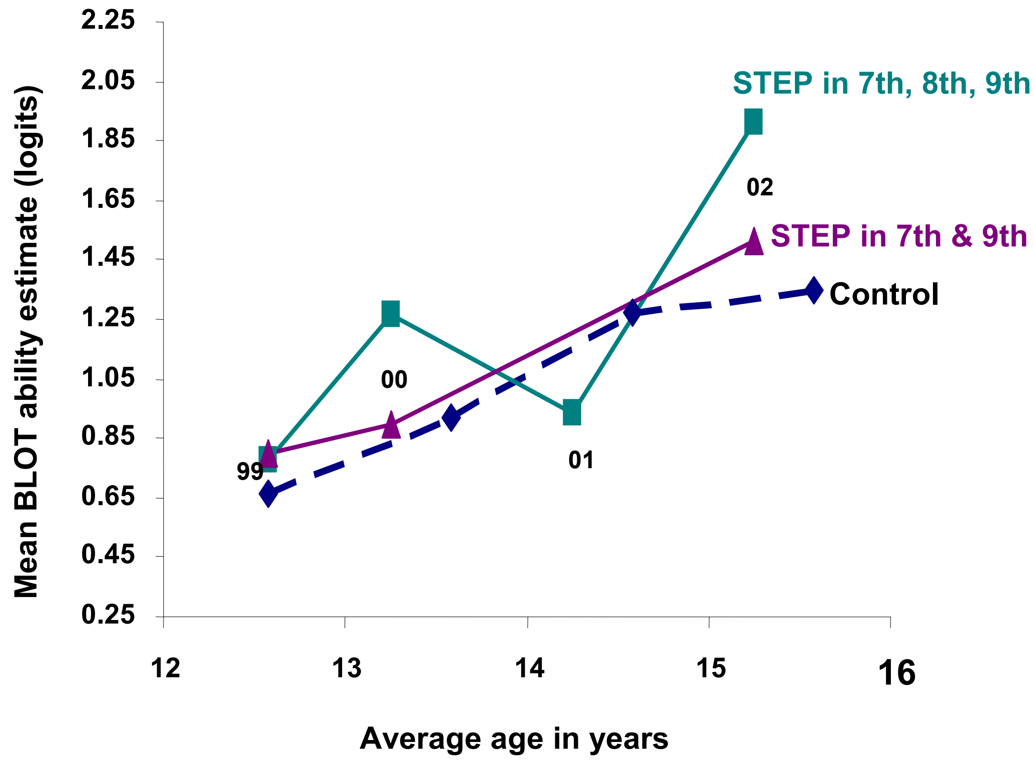


Figure 5.15 Peers of Cohort B who missed STEP in 8th grade

Table 5.17
Mean BLOT ability estimates of peers of Cohort A who missed STEP in 6th grade

		Mean BLOT ability estimate (logits)		
Grade	**Average age (yrs; mths)**	**Control**	**STEP in 6th, 7th & 8th grade**	**STEP in 7th & 8th grade only**
6th	**11; 6**	-0.34 (50)	-0.05 (46)	
6th	**12; 3**		0.50 (82)	
7th	**12; 6**	0.66 (166)		
7th	**13; 3**		0.99 (80)	0.9 (24)
8th	**13; 6**	0.92 (140)		
8th	**14;3**		1.43 (46)	1.42 (24)
9th	**14;6**	1.27 (175)		

Table 5.18
Mean BLOT ability estimates of peers of Cohort B who missed STEP in 8th grade

		Mean BLOT ability estimate (logits)		
Grade	**Average age (years; months)**	**Control**	**STEP in 7th, 8th & 9th grade**	**STEP in 7th & 9th grade only**
7th	**12; 6**	0.66 (166)	0.78 (161)	0.8 (17)
7th	**13; 3**		1.26 (171)	0.9 (17)
8th	**13; 6**	0.92 (140)		
8th	**14; 3**		0.93 (139)	
9th	**14; 6**	1.27 (175)		
9th	**15; 3**		1.91 (103)	1.51(17)
10th	**15; 6**	1.35 (91)		

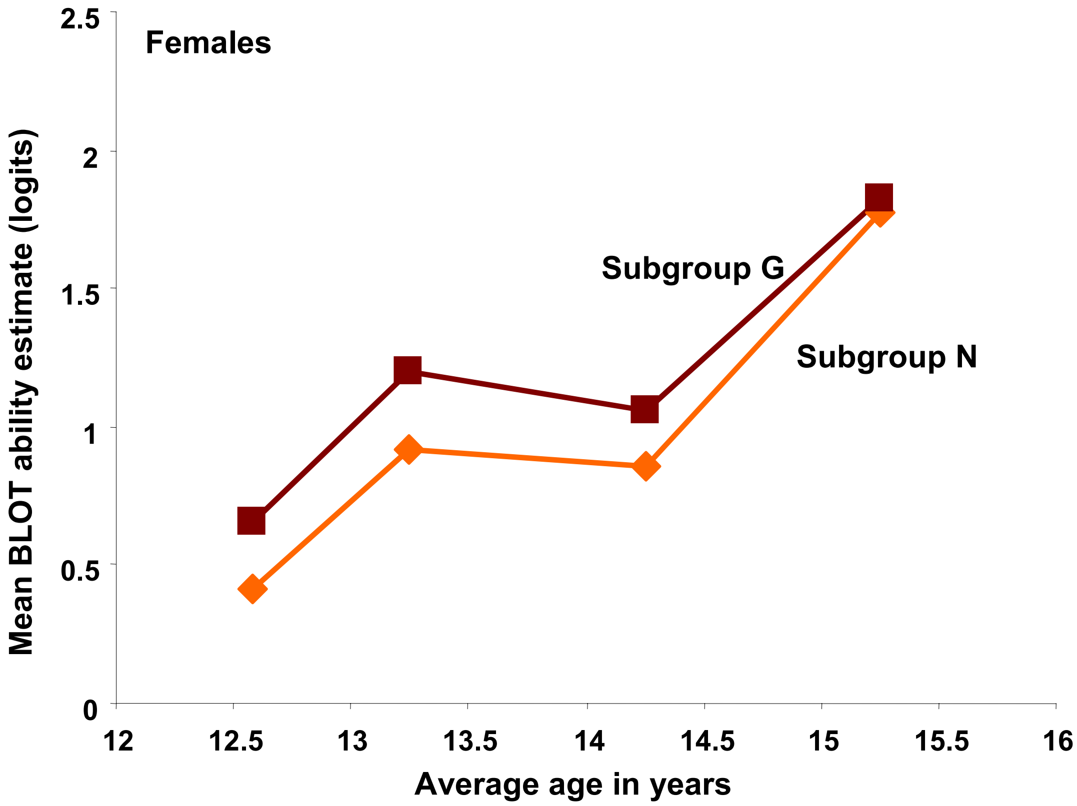
A second subset of students was identified that experienced only part of the STEP intervention. These were peers of Cohort B whose 8th grade teachers opted out of the STEP in the 2000-2001 school year. This subset of students was tested with the BLOT in September 1999, May 2000 and then in May 2002. Their results can be seen in Table 5.18

and Figure 5.15. In one-sample, one-tailed *t*-tests, no significant difference was found between the subset of students who missed the STEP in 8th grade and either the pre-intervention control (*t* [df], *p*: 0.98 [21, NS]), or their peers in Cohort B (*t* [df], *p*: 1.35 [31, NS]). These results suggest that the portion of the intervention they did receive might have been too little, and too late in their adolescence, to affect their cognitive development. This is an important result, that draws attention to critical aspects associated with the delivery of the CASE strategies.

The students who did not experience the STEP intervention when they were in the 8th grade in 2000-2001 missed a number of CASE lessons (see Table 4.5, page 69), that reduced their total of CASE activities to 12, compared with the 21 or 16 activities experienced by their peers (dependent on which teacher they had in 8th grade). A one sample *t*-test was performed to compare the mean change in the BLOT ability estimate for 2001 to 2002 of students in Cohort B, who experienced all three years of the STEP intervention, with that of an interpolated mean change for their peers who missed the STEP in 8th grade. A significant difference was found (*t*[df] = 3.08 [33], *p* = <0.01) for the last year of the intervention, with a large effect size of 1.07 standard deviations. This important result suggests that missing a year of the STEP intervention had a strongly negative effect on the cognitive development of these peers of Cohort B.

It is difficult to explain the difference in cognitive change from 1999 to 2000 between Cohort B and their peers who later missed the STEP in 8th grade (see Figure 5.15). The students started at a similar cognitive developmental level, but those who went on to complete the whole intervention were more developed by May 2000. It is not known whether these two subsets of students received the same treatment during the first year of the intervention, as the BLOT answer sheets did not identify the teacher. The 7th grade teachers all claimed to have completed the first five CASE activities, but there was no means of verifying this, nor whether the CASE lessons had been taught in a similar manner. It is possible that some factor associated with the delivery of the STEP program caused this surprising difference between these two groups of students. Unfortunately the Molalla River Middle School teachers did not keep their annual grade books and no records exist of the class lists of that year. It might be that age 12+ to 13+ is a critical age, when some students are primed to accelerate in their cognitive development given an optimal delivery of the STEP intervention.

The two 8th grade science teachers (pseudonyms Nancy and George) delivered different programs of CASE activities to Cohort B in 2000-2001 (see Table 4.5, page 69). It was possible to identify which students were taught by which teacher from data sheets of the 8th grade Oregon State Science Tests. Students in Cohort B who were taught by Nancy (subgroup N) covered five more CASE activities than those taught by George (subgroup G). When the change in the BLOT ability estimate of these two subsets of STEP students was compared in a two-sample *t*-test, no significant difference was found for either of the last two years of the intervention (2000-2001: t [df], p = -0.02 [88], NS; 2001-2002: t [df], p = 0.50 [81], NS). The effect sizes were negligible, -0.01 and 0.11 SDs respectively. It appears that the different treatment that these subsets of students received in 8th grade did not affect the considerable growth spurt that many students experienced in the last year of the intervention (see Figure 5.16).



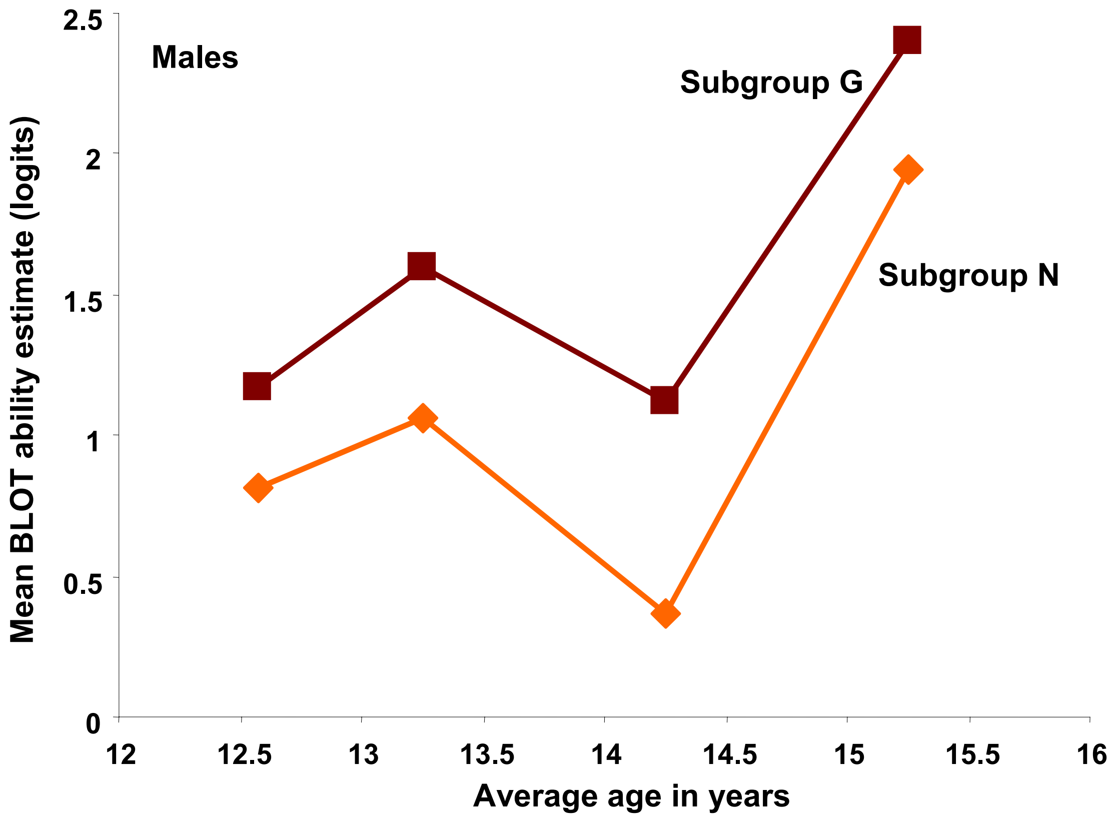


Figure 5.16 Cognitive development profiles of students in subgroups of Cohort B. Subgroup N was taught by Nancy, subgroup G by George.

Analysis by stage of cognitive development

Before the cognitive tests (Piagetian Reasoning Tasks) used in the original CASE work and the BLOT test were shown to have exemplary Rasch scale measurement characteristics, authors in this genre (Adey & Shayer, 1994; Endler & Bond, 2002) converted

raw score to Piagetian level of cognitive development in order to examine the effect of CASE interventions on the cognitive level of children. These Piagetian cognitive levels are routinely known as IIA (early concrete), IIB (mature concrete), IIIA (early formal), and IIIB (mature formal). Subsequently, Rasch analysis (based on the principle that the number of totally correct answers is the sufficient summary statistic for ability) confirmed the validity of converting raw score to Rasch scale and raw score to Piagetian level (see Table 5.19).

Table 5.19

Thresholds for Rasch ability estimate, BLOT raw score and Piagetian level

Rasch ability estimate (logits)	BLOT raw score	Piagetian level
2.58	32	Mature formal (IIIB)
1.74	29	Early formal (IIIA)
0.18	19	Mature concrete (IIB)
< 0.18	< 19	Early concrete (IIA)

The raw scores of the 1999 BLOT pre-intervention data were used to estimate the cognitive level of each student in the Molalla River pre-intervention population. Table 5.20 and Figure 5.17 show the estimated cognitive levels of the student body of the Molalla River School District in September 1999, prior to the commencement of the STEP intervention.

Figure 5.17 shows in cross-section the overall progression from concrete (IIA/B) to formal (IIIA/B) operational thinking through the secondary school years. The majority of students in 6th grade are at the early concrete (IIA) level, with few mature formal thinkers. Whereas by 10th grade, the majority of students have attained the mature concrete (IIB) or early formal (IIIA) level. It is interesting to note that there are relatively few mature formal (IIIB) operational thinkers in the 10th grade population (13.33%), with little change from the 9th grade percentage (13.14%). This similarity between the 9th grade (mean age 15+) and 10th grade (16+) data is comparable with the results of Shayer and Wylam (1978), who reported no increase in the proportion of formal operational thinking beyond age 15 years in a sample of 1,200 British children.

Table 5.20
Estimated cognitive level of the pre-intervention control population

	% of students					
Cognitive level	**6th grade**	**7th grade**	**8th grade**	**9th grade**	**10th grade**	**11th grade**
IIA early concrete	67.31	36.00	27.70	18.29	17.78	12.00
IIB mature concrete	21.15	47.33	50.68	46.86	43.33	25.00
IIIA early formal	9.62	10.67	15.54	15.54	25.56	25.56
IIIB mature formal	1.92	6.00	6.08	13.14	13.33	8.00

The cognitive levels of the STEP students in May 2002 were similarly estimated from the BLOT raw scores and these are shown in Table 5.21. A comparison of the estimated cognitive levels of the pre-intervention control and the STEP students in May 2002 can be seen in Figure 5.18. The very low probabilities calculated from the Fisher's Exact test (see Table5.22) indicate that there are significant differences between the cognitive levels of the control and STEP students. This, in combination with the direction of the deviation, suggests that students who had experienced the STEP intervention were indeed more cognitively developed than their age peers in the Molalla River School District prior to the commencement of the STEP intervention.

Table 5.21
Estimated cognitive levels of STEP students in May 2002

	% of students		
Cognitive level	**8th grade**	**9th grade**	**10th grade**
IIA early concrete	26.5	10.7	17.9
IIB mature concrete	20.6	35.9	28.2
IIIA early formal	23.5	23.3	15.4
IIIB mature formal	29.4	30.1	38.5

Table 5.22
A comparison of the estimated cognitive levels of the control and STEP students (May 2002)

		Number of students				
Grade		**IIIB**	**IIIA**	**IIB**	**IIA**	***P***
8th	**Control**	9	24	69	40	0.00011
	STEP	8	10	8	9	
9th	**Control**	23	39	81	32	0.000089
	STEP	27	28	38	10	
10th	**Control**	12	23	37	18	0.0082
	STEP	11	10	11	7	

(Probabilities calculated by Fisher's Exact Test)

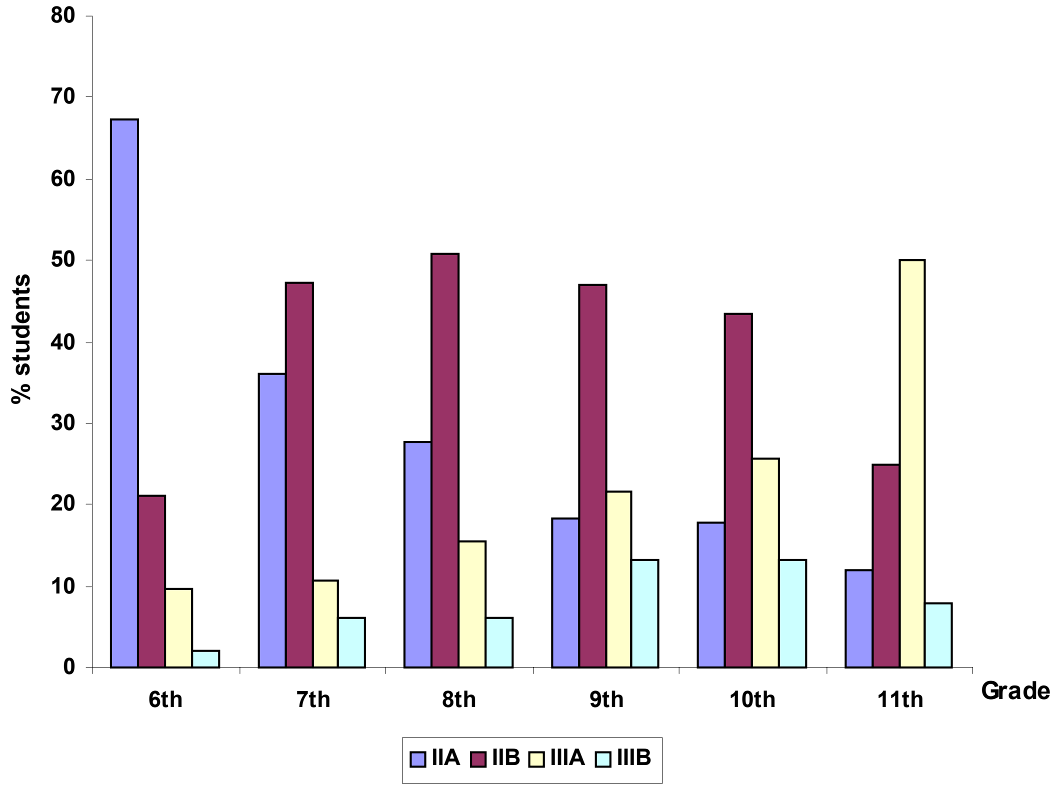


Figure 5.17 Estimated cognitive levels of the control sample (BLOT 09/1999)

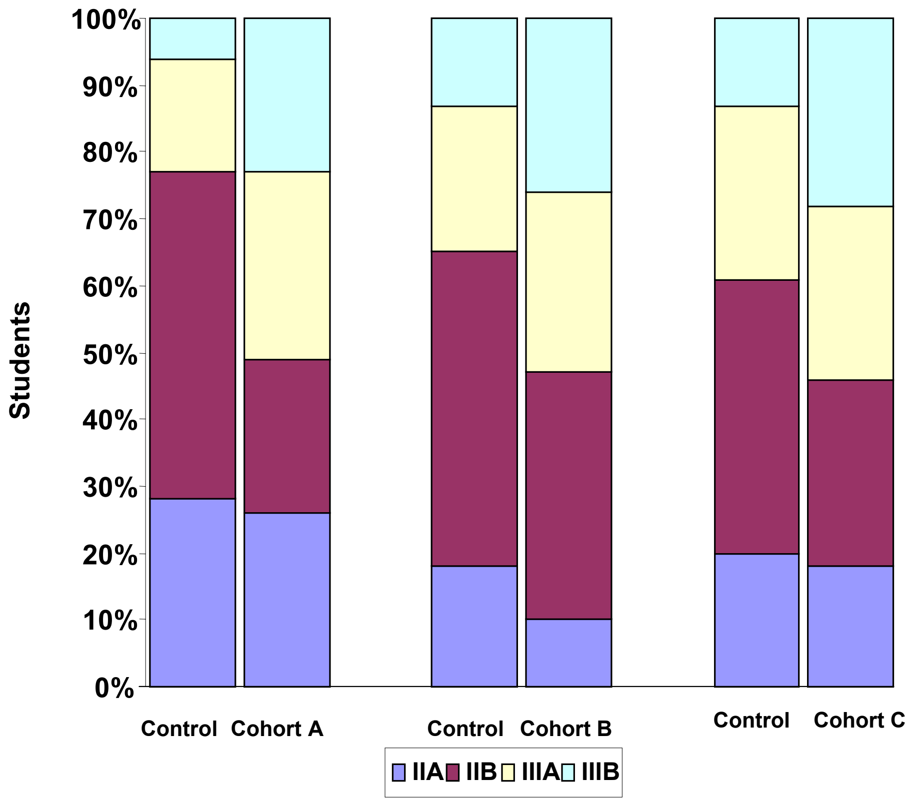


Figure 5.18 Estimated cognitive levels of the control sample and STEP students (05/02)

CHAPTER 6

Student achievement

Student achievement in Mathematics and Reading & Literature is tested in the spring at all grade levels in Oregon. Science, however, is tested at 5^{th}, 8^{th} and 10^{th} grades only. It follows that a more detailed picture of the student achievement of the STEP students can be obtained from an analysis of the STEP students' Mathematics and Reading & Literature Oregon State Scores than from their Science results. Students in Cohorts A and B experienced only one state-mandated Science test (at 8th grade) during the period of this study, whereas Cohort C was tested at both 8th and 10th grades. School-based achievement tests were considered to be too class-specific in nature to be of value in following the science achievement of the students over time.

Master data file

The master file of results for student achievement was based on the total number of students who had completed the STEP intervention available to their grade, and had taken the relevant state-mandated test in the spring of 1999 and again on at least one other subsequent occasion. In order to make a thorough examination of the possible effects of the STEP intervention on student achievement, a number of data analytical perspectives were taken. For every comparison, the maximum complete data set (*i.e.* all those students without relevant missing data cells) was included in the calculation. For the estimation of change over time, and the subsequent comparison with the control data, students included in the growth estimate were the children in the relative cohort who were tested at both the beginning and the end of the appropriate annual period. Therefore, the sample sizes for the growth comparisons (students tested twice in twelve months) seen, for example, in Table 6.5, are necessarily smaller than those in Table 6.2, that shows the mean results for children tested merely twice in the 36-month period.

The Mathematics achievement of STEP students

In the absence of control data from a relevant non-project cohort, it was necessary to create a pre-intervention cross-sectional control profile of Mathematics achievement. The data came from the Mathematics Oregon State Scores (OSS) for the 5th to 9th grade state-mandated tests of May 1999 (4 months prior to the start of the STEP intervention). The pattern shown by the cross-sectional control profile is one of little change in Mathematics OSS from mean age 10 + to 11+, a growth phase from mean age 11+ to 13+ years, followed by a period of slower growth (see Table 6.1, Figure 6.1a). The Oregon Department of Education benchmark scores for meeting and exceeding the 5th, 8th and 10th grade standards are also displayed in Figure 6.1. The control sample met the 5th grade benchmark for Mathematics, but not the 8th grade.

Comparisons were made between the control profile for Mathematics achievement and those of the three cohorts of STEP students that were tested in May 1999, 2000, 2001 and 2002 (Table 6.2 and Figure 6.1b, c, d). Figure 6.1b shows that Cohort A met the 5th and 8th grade Oregon benchmark. Cohort B exceeded the 8th grade benchmark (Figure 6.1c), and Cohort C students met the 8th, but not the 10th grade benchmark (Figure 6.1d).

Table 6.1
Mean Mathematics OSS of control students (1999)

Grade (N students)	**Average age (Yrs; mths)**	**Mean Mathematics OSS (RITs)**
5th (43)	10; 3	218.77
6th (198)	11; 3	217.59
7th (188)	12; 3	221.42
8th (223)	13; 3	229.02
9th (185)	14; 3	230.01
10th (174)	15; 3	233.95

Table 6.2
Mean Mathematics OSS of STEP students (1999, 2000, 2001, 2002)

Grade	**Average age**	**Mean Mathematics OSS in RITs (N students)**			
	(Yrs; mths)	**Control**	**Cohort A**	**Cohort B**	**Cohort C**
5th	**10; 3**	218.77 (43)	221.40 (46)		
6th	**11; 3**	217.59 (198)	219.50 (50)	220.00 (84)	
7th	**12; 3**	221.42 (188)	226.24 (50)	228.20 (97)	233.60 (56)
8th	**13; 3**	229.02 (223)	235.38 (50)	240.12 (100)	231.70 (61)
9th	**14; 3**	230.01(185)		233.39 (101)	230.12 (60)
10th	**15; 3**	233.1(174)			234.11 (45)

Regression analyses were performed on the pre-intervention (05/1999) and post-intervention (05/2002) Mathematics OSS. The results indicated that significant improvement had taken place in all three cohorts of STEP students over the period of the intervention (t[df], p, Cohort A: 7.25 [44], <0.001; Cohort B: 11.89 [78], <0.001; Cohort C: 7.50 [39], 0.001). The effect sizes of these results are all large (> 2.00 SDs).

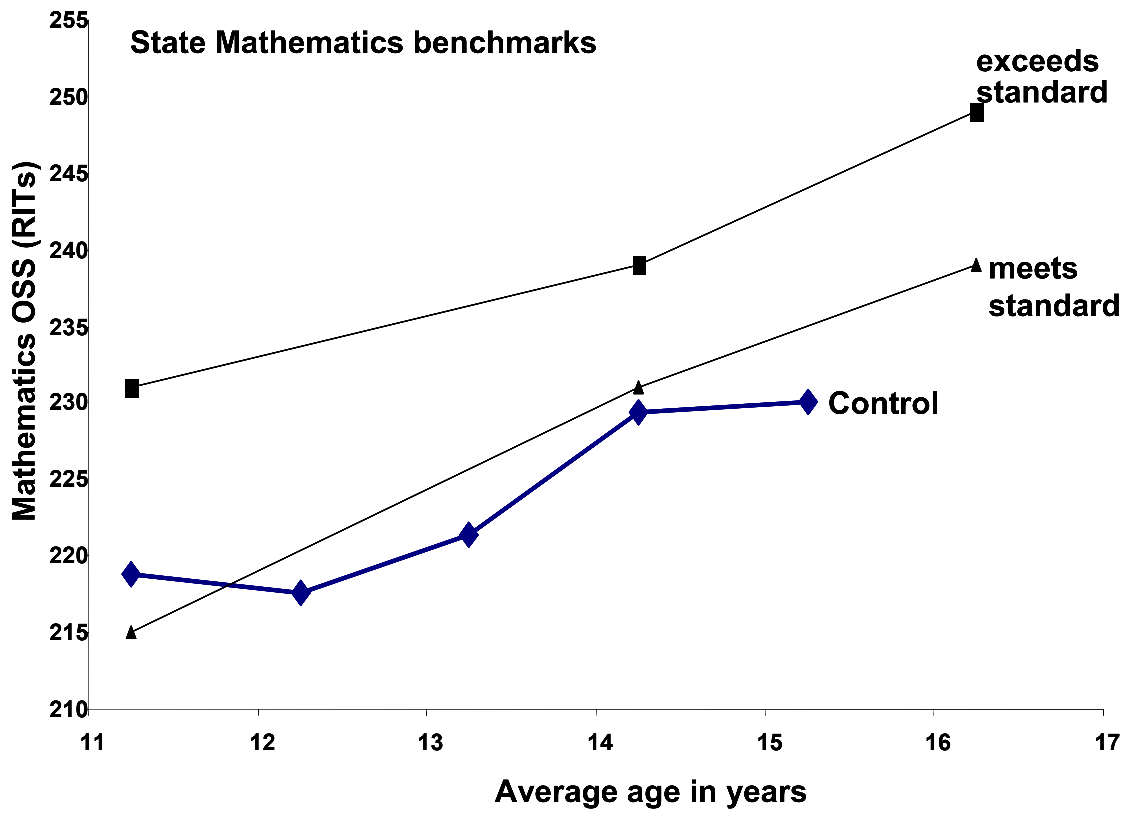


Figure 6.1a Mean Mathematics OSS of control

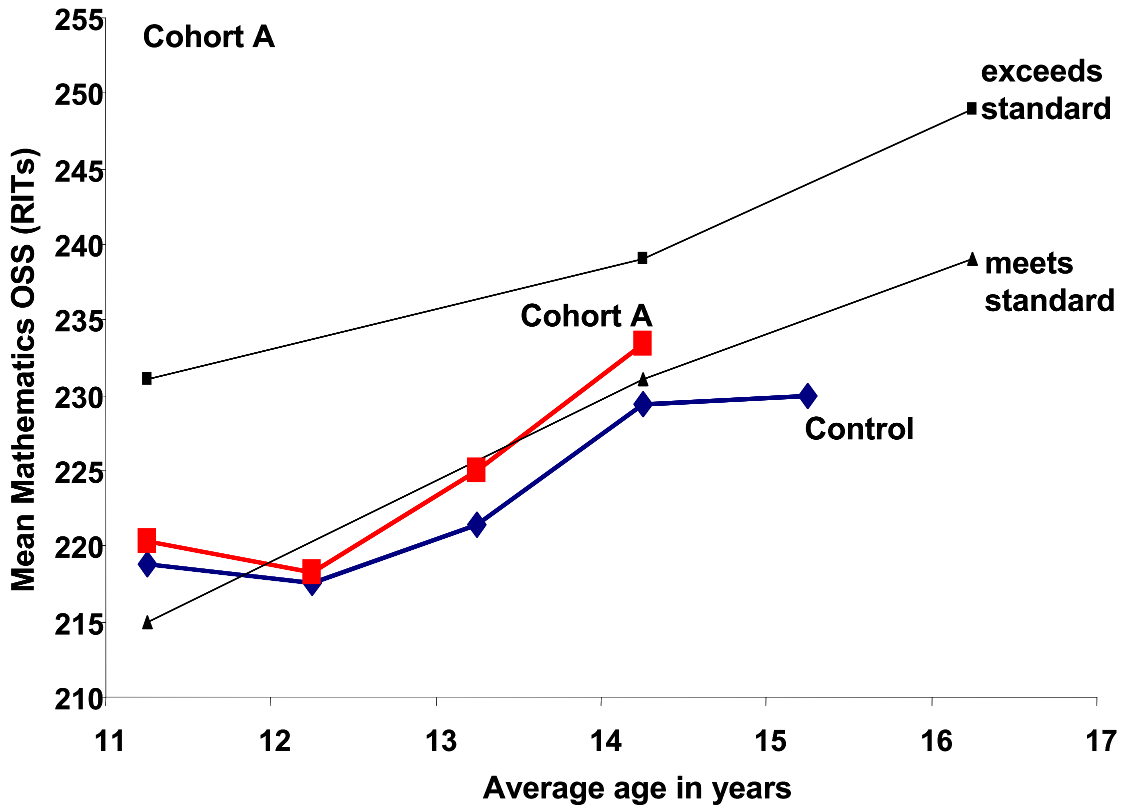


Figure 6.1b Mean Mathematics OSS of Cohort A

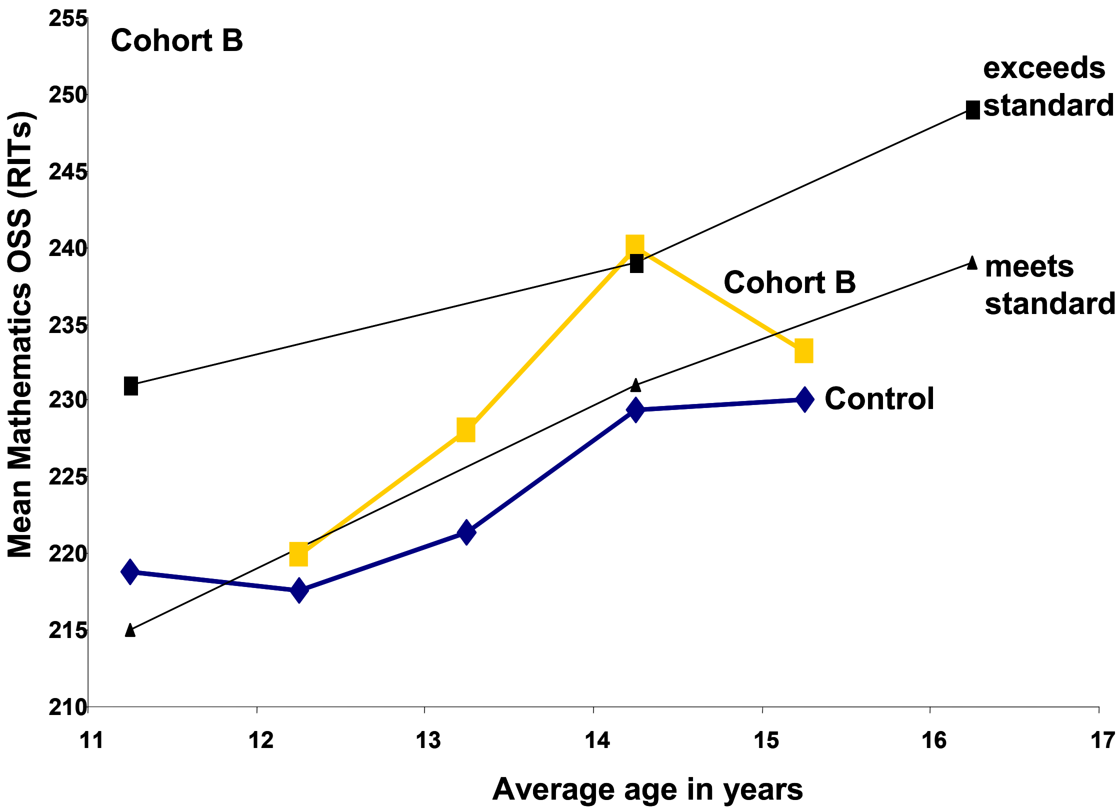


Figure 6.1c Mean Mathematics OSS of Cohort B

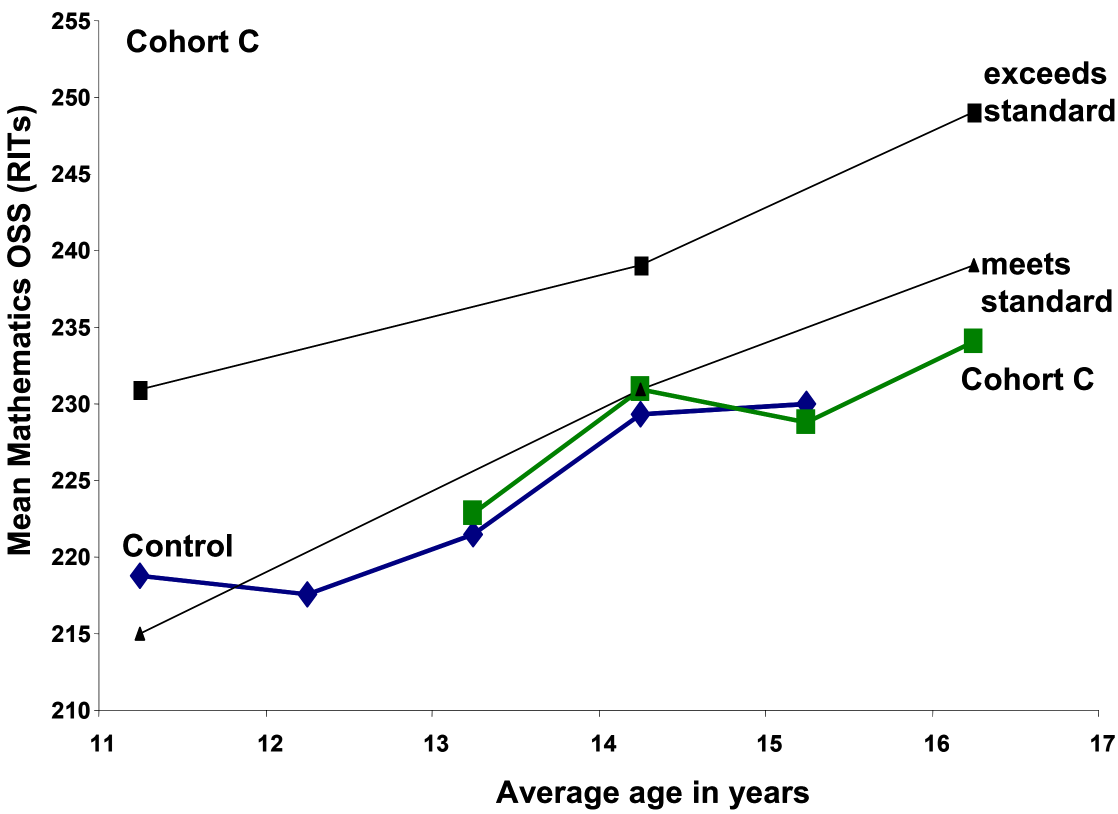


Figure 6.1d Mean Mathematics OSS of Cohort C

Overall slopes and annual changes in the Mathematics OSS for the control sample were calculated from the means of the pre-intervention population and are displayed in Table 6.3. The overall slope and annual changes in Mathematics OSS for each individual STEP student were also calculated, as described for the BLOT analysis in Chapter 5. The means of these overall slopes and annual changes for each STEP cohort are shown in Table 6.4. The Mathematics OSS control slopes were used as parameters in *t*-tests to compare overall change over time for the control and STEP students. As with the BLOT analysis, periodic change was also investigated, using the control mean change as the parameter. One-tailed probabilities were used in the *t*- tests because there is an *a priori* expectation that the treatment has a positive effect (Adey & Shayer, 1994).

Table 6.3
Mean change in Mathematics OSS of the control

	Mean change in Mathematics OSS in RITs			
Cohort equivalent	**Overall slope (change in OSS/time in months)**	**0-12 months**	**12-24 months**	**24-36 months**
A	0.47	-1.18	3.84	7.60
B	0.28	3.84	7.60	0.98
C	0.07	7.60	0.98	3.95

Table 6.4
Mean change in Mathematics OSS of STEP students (1999 to 2002)

	Mean change in Mathematics OSS in RITs (N students)			
Cohort	**Overall slope (change in OSS/time in months)**	**1999-2000**	**2000-2001**	**2001-2002**
A	0.44 (50)	-1.84 (45)	6.86 (49)	9.16 (50)
B	0.39 (101)	7.54 (81)	12.21 (97)	-6.71 (100)
C	0.24 (62)	7.91 (57)	-1.63 (60)	4.79 (43)

Table 6.5
Comparison of mean change in Mathematics OSS between control and STEP students

Cohort		Mean change in Mathematics OSS in RITs (N students) Overall (99-02)	05/99-05/00	05/00-05/01	05/01-05/02
A	***t*(df)**	4.17 (50)	-0.55 (44)	3.11 (48)	1.58 (49)
	p	0.0001*	0.29 NS	0.0016*	0.060 NS
	Effect size *d*	1.18, large	-0.17 (nil)	0.90, large	0.45 (small)
B	***t*(df)**	2.04 (100)	4.96 (80)	6.95 (96)	-11.11 (99)
	p	0.022*	<0.0001*	<0.0001*	<0.0001*
	Effect size *d*	0.41, small	1.11, large	1.42, large	-2.23, large
C	***t*(df)**	0.59 (61)	0.92 (56)	-2.61 (59)	0.98 (42)
	p	0.28 NS	0.18 NS	0.0057*	0.17 NS
	Effect size *d*	0.15 (nil)	0.25 (small)	-0.68, medium	0.30 (small)

The *t*-tests are one-sample tests comparing control and experimental slopes (column1) and annual change (columns 2, 3, 4). ***t*** represents the *t* statistic, **df** represents degrees of freedom, and ***p*** represents probability of the null hypothesis. * indicates significance with the sequential Bonferroni test (∝ = 0.05).

Significant overall gains from 1999 to 2002 were made by Cohorts A (1.18 SDs) and B (0.41 SDs). Cohort A also showed a statistically significant large gain over the control in the second year (0.90 SDs). Cohort B had significant gains in the first two years (1.11, 1.42 SDs), followed by significantly lower (-2.23 SDs) results in 2002. No statistically significant gains over the control were apparent for Cohort C, however, there was a significant loss (-0.68 SDs) from 2000-2001.

The Chair of the Mathematics Department at Molalla River High School claimed to be unable to offer an explanation for the mean Mathematics results of Cohort B being lower in May 2002 than in May 2001 (personal communication, 2003). The fall in mean scores was also evident, though to a lesser extent, in the Reading & Literature scores of the same students. These interesting results will be discussed further in Chapter 9.

Regression analysis was performed on the level of parent education and change in Mathematics OSS from 1999 to 2002 (t[N], p = 1.25 [132], NS). No significant relationship was found, which suggests that the Mathematics achievement of these students is not associated with the educational level of their parents (see Figure 6.2). When a regression analysis was performed on the starting level (Mathematics OSS 1999) and the change in Mathematics OSS from 1999 to 2002, a significant relationship was found (t[N], p: -4.36 [165], <0.001), suggesting that a lower starting level is associated with a greater gain in Mathematics achievement for the STEP cohorts (see Figure 6.3). However, when a regression was performed of the age when the STEP intervention commenced and the change in Mathematics OSS (see Figure 6.4), the result was not significant (t[N], p = -1.66 [167], NS). These results suggest that improvement in the Mathematics achievement of STEP students is more likely to be associated with starting level than with age when the intervention commenced. This is an important result, because it shows that relatively high gains in Mathematics can be made by some students across the spectrum of age when the intervention commenced.

The results for Mathematics achievement were analysed separately for male and female students. The mean Mathematics OSS for the males and females of each cohort are shown in Table 6.6, along with the means of the control profile. Comparison of the male and female control students revealed minor differences in the patternof Mathematics achievement for the sexes (see Figure 6.5a). Figure 6.5a shows a growth spurt for females at age 10+, whereas for males it is at 11+.

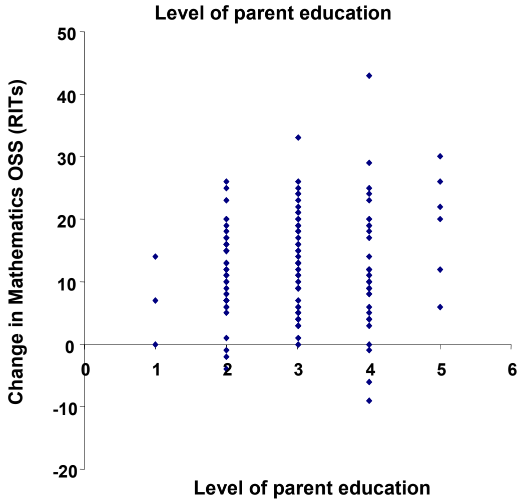


Figure 6.2 Level of parent education and change in `
Mathematics OSS

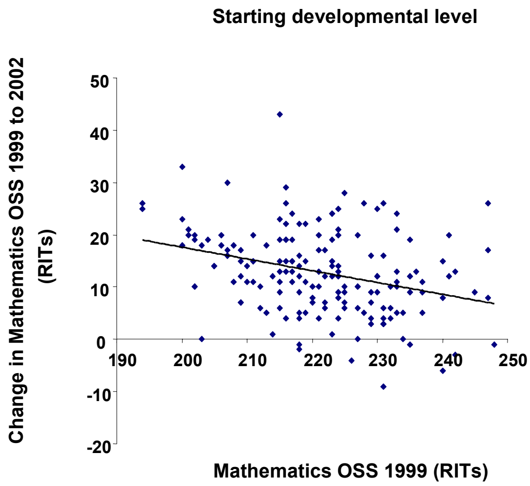


Figure 6.3 Starting developmental level and change in
Mathematics OSS

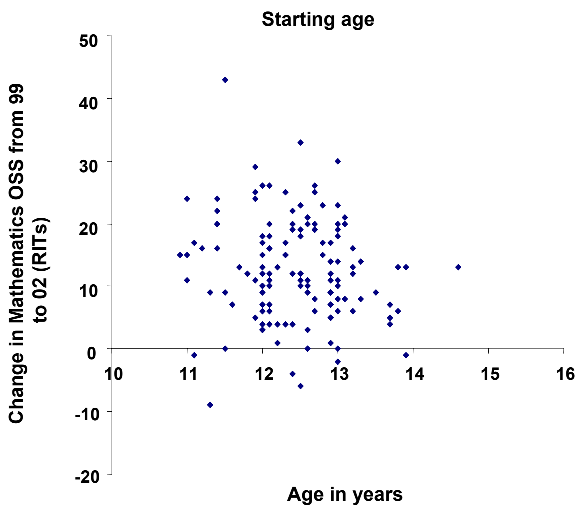


Figure 6.4 Starting age and change in Mathematics
OSS

Table 6.6
Mean Mathematics OSS of male and female students

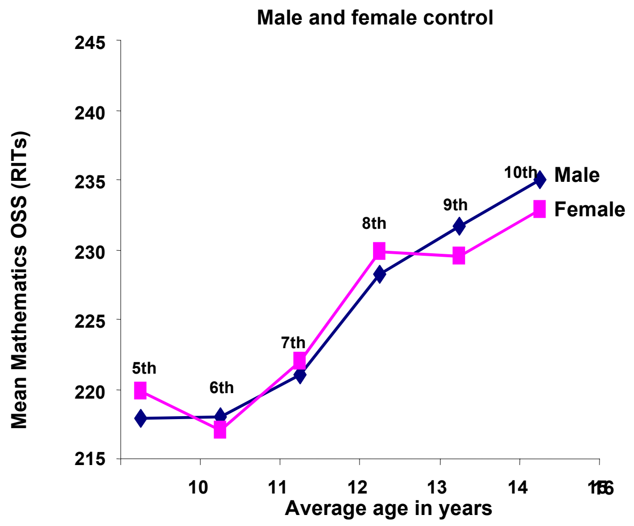
			Mean Mathematics OSS in RITs (N students)			
Grade	**Average age (Yr;mth)**	**Sex**	**Control**	**Cohort A**	**Cohort B**	**Cohort C**
5th	**10; 3**	m	217.92 (24)	221.30 (24)		
		f	219.80 (19)	221.50 (22)		
6th	**11; 3**	m	218.00 (91)	219.67 (27)	226.00 (27) *	
		f	217.00 (107)	219.30 (23)	217.20 (57)	
7th	**12; 3**	m	221.00 (96)	225.89 (27)	232.61 (33) *	223.10 (27)
		f	222.00 (92)	226.65 (23)	225.92 (64)	224.20 (29)
8th	**13; 3**	m	228.21 (114)	234.85 (27)	244.66 (35) *	231.35 (31)
		f	229.87 (109)	236.00 (23)	237.68 (65)	232.07 (30)
9th	**14; 3**	m	231.72 (39)		237.31 (35) *	229.31 (32)
		f	229.55 (146)		231.30 (66)	231.04 (28)
10th	**15; 3**	m	235.00 (88)			234.85 (26)
		f	232.88 (66)			233.11 (19)

* indicates a significant difference between males and females.

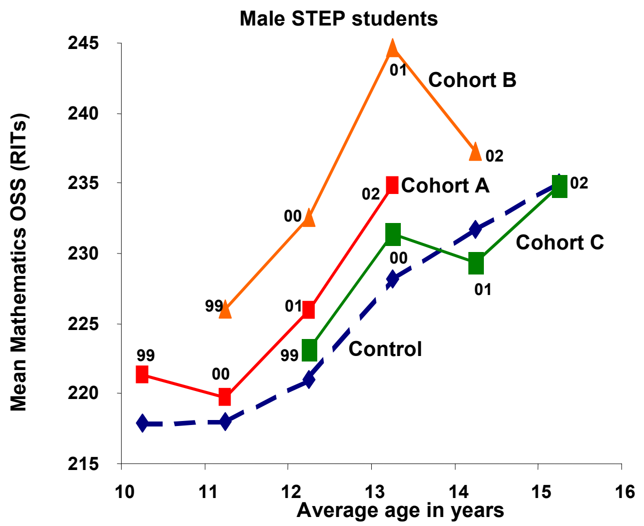
No significant difference was found between the mean Mathematics OSS of males and females in Cohorts A and C. In contrast, males significantly outperformed females in Cohort B (> 0.8 SDs).

The results of one-tailed *t*-tests comparing the slopes of same sex control and STEP students are shown in Tables 6.7 and 6.8. One-tailed probabilities were used in these *t*-tests because of the *a priori* expectation that the treatment has a positive effect (Adey & Shayer, 1990). A significant overall gain (1.50 SDs, large) was found for the slope of the females in Cohort A over the control. No statistically significant overall effect was evident for Cohorts B and C females. However, there were large gains over the control for the females in Cohort B in the first and second year respectively (0.99 and 1.16 SDs),

a)



b)



c)

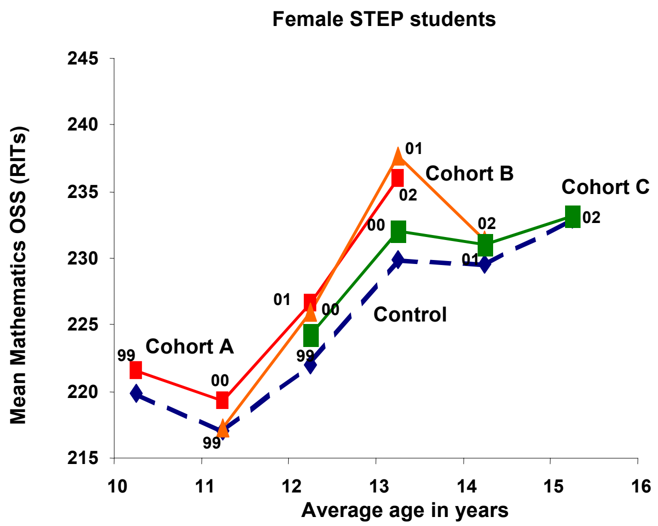


Figure 6.5 Mean Mathematics OSS of males and females

with a significant loss (-1.76 SDs) in the third year of the intervention.

Table 6.7
Comparison of mean change in Mathematics OSS between female control and STEP students

Cohort		Mean change in Mathematics OSS of females in RITs (N students) Overall (99-02)	05/99-05/00	05/00-05/01	05/01-05/02
A	***t*(df)**	3.52 (22)	0.58 (20)	2.39 (21)	1.00 (22)
	p	0.00096*	0.28 NS	0.013 NS	0.16 NS
	Effect size *d*	1.50, large	0.26 (small)	1.04 (large)	0.43 (small)
B	***t*(df)**	1.50 (65)	3.59 (53)	4.56 (62)	-7.06 (64)
	p	0.069 NS	0.00036*	<0.0001*	<0.0001*
	Effect size *d*	0.37 (small)	0.99, large	1.16, large	-1.76, large
C	***t*(df)**	-2.04 (30)	-0.56 (28)	-0.77 (27)	1.05 (16)
	p	0.025 NS	0.29 NS	0.22 NS	0.15 NS
	Effect size *d*	-0.74 (medium)	-0.21 (small)	-0.30 (small)	0.53 (medium)

The *t*-tests are one-sample tests comparing control and experimental slopes (column1) and annual change (columns 2, 3, 4). ***t*** represents the *t* statistic, **df** represents degrees of freedom, and ***p*** represents probability of the null hypothesis. * indicates significance with the sequential Bonferroni test (∝ = 0.05).

Overall, only males in Cohort A made a significant gain (1.08 SDs, large) over the control. However, Cohort B males showed statistically significant gains in the first (0.94 SDs) and second year (1.19 Sds) of the intervention, followed by a significant loss (-3.13 SDs) in 2001-2002. No statistically significant effects were seen for Cohort C over the control.

A comparison of the Mathematics achievement of males and females for each cohort is shown in Figure 6.6. The patterns of Mathematics achievement of males and females in Cohort A are revealed as very similar, and no significant difference was found in any of the annual slopes of change. Similar patterns of Mathematics achievement are also evident for Cohorts B and C, although Cohort B males consistently scored 5 to 10 RITs higher than their female peers.

Table 6.8
Comparison of mean change in Mathematics OSS between male control and STEP students

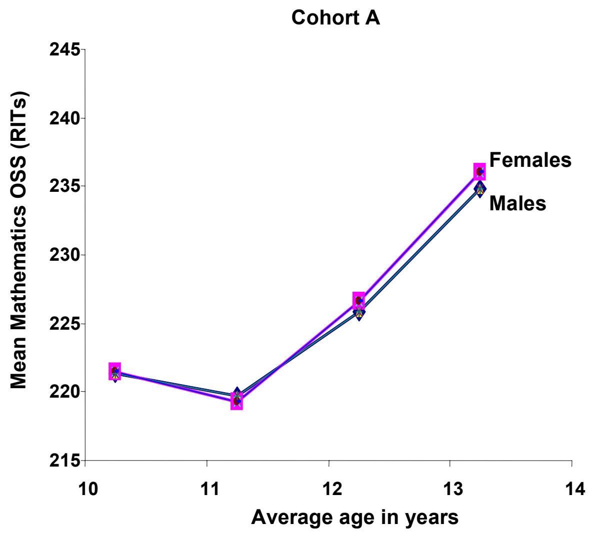
Cohort		**Mean change in Mathematics OSS of males in RITs (N students)** Overall (99-02)	05/99-05/00	05/00-05/01	05/01-05/02
A	***t*(df)**	2.81 (27)	-1.19 (23)	2.12 (23)	1.33 (26)
	p	0.0046*	0.12 NS	0.022 NS	0.097 NS
	Effect size *d*	1.08, large	0.50 (medium)	0.88 (large)	0.52 (medium)
B	***t*(df)**	-0.08 (34)	2.39 (26)	5.49 (33)	-9.12 (34)
	p	0.47 NS	0.01*	<0.0001*	<0.0001*
	Effect size *d*	-0.03 (nil)	0.94, large	1.19, large	-3.13, large
C	***t*(df)**	0.54 (31)	1.05 (62)	-0.71 (65)	1.19 (47)
	p	0.30 NS	0.15 NS	0.24 NS	0.12 NS
	Effect size *d*	0.19 (nil)	0.26 (small)	-0.18 (nil)	0.35 (small)

The *t*-tests are one-sample tests comparing control and experimental slopes (column1) and annual change (columns 2, 3, 4). ***t*** represents the *t* statistic, **df** represents degrees of freedom, and ***p*** represents probability of the null hypothesis. * indicates significance with the sequential Bonferroni test (∝ = 0.05).

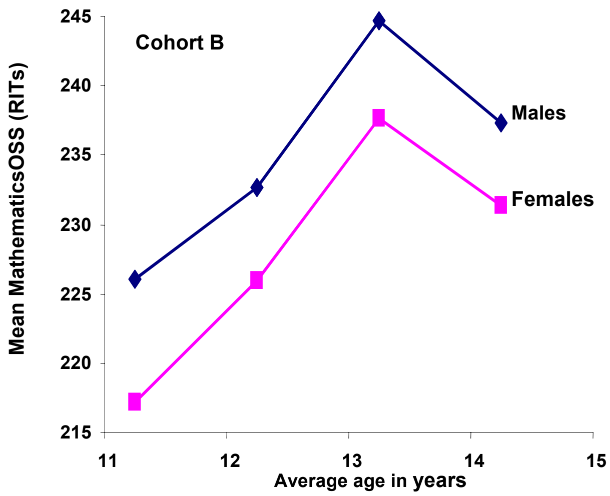
Mathematics results for students who experienced only part of the STEP intervention

The Mathematics results were analysed for those peers of Cohort A who did not experience the STEP intervention at 6th grade in 1999-2000 (see Table 6.9 and Figure 6.7). Some of these students, who began the STEP at 7th grade (September 2000), wer new to the school, whereas others had been in classes that did not participate in STEP in 6th grade.

a)



b)



c)

Figure 6.6 Mean Mathematics OSS of males and females in STEP Cohorts

Table 6.9
Mean Mathematics OSS of Cohort A students who began STEP in 7th grade

		Mean Mathematics OSS in RITs (N students)		
Grade	**Average age** (Yr; mth)	**Control**	**STEP in 6th, 7th & 8th grade**	**STEP in 7th & 8th grade only**
5th	**10; 3**	218.77 (43)	221.40 (46)	224.33 (18)
6th	**11; 3**	217.59 (198)	219.50 (50)	222.61 (18)
7th	**12; 3**	221.42 (188)	226.24 (50)	228.59 (22)
8th	**13; 3**	229.02 (223)	235.38 (50)	234.95 (19)

The latecomers started out at a higher mean level than the students who had completed one year of the STEP intervention. However, Figure 6.7 suggests that Cohort A closed the gap and overtook the latecomers by the end of 2002. A one-tailed *t*-test was performed to compare the change in Mathematics OSS of the latecomers to the STEP with the pre-intervention control population. The latecomers showed a significant difference in the last year of the intervention compared with the control (*t* [df], *p*: 2.56 [21], <0.02). The large effect size (1.12 SDs) might indicate that the STEP intervention had enhanced the Mathematics achievement of these students. The use of this one-tailed test can be justified by the *a priori* assumption that the CASE strategies enhance Mathematics achievement (Adey & Shayer, 1994).

A two-tailed *t*-test of the difference between the means of the late comers and their STEP peers indicated that there was a significant difference between the means of the slopes in the third year of the intervention ((*t*[df], *p*: 2.20 [29], <0.05). The effect size was large (0.82 SDs), suggesting that the effect of three years of the STEP on Mathematics achievement was more beneficial than two years for these students. This result is

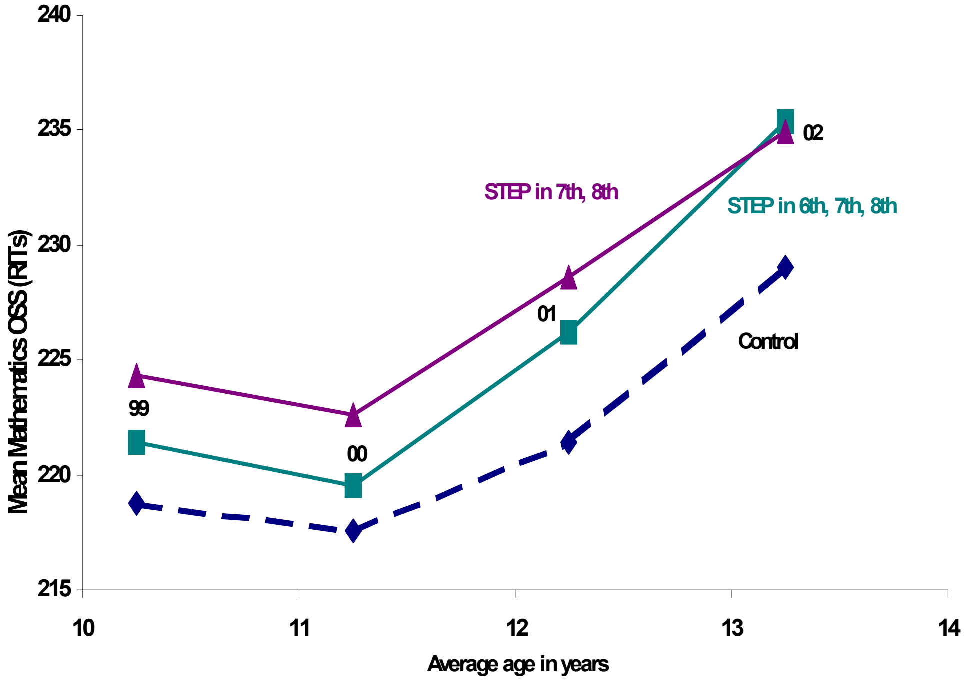


Figure 6.7 Mean Mathematics OSS of peers of Cohort A who missed STEP in 6th grade

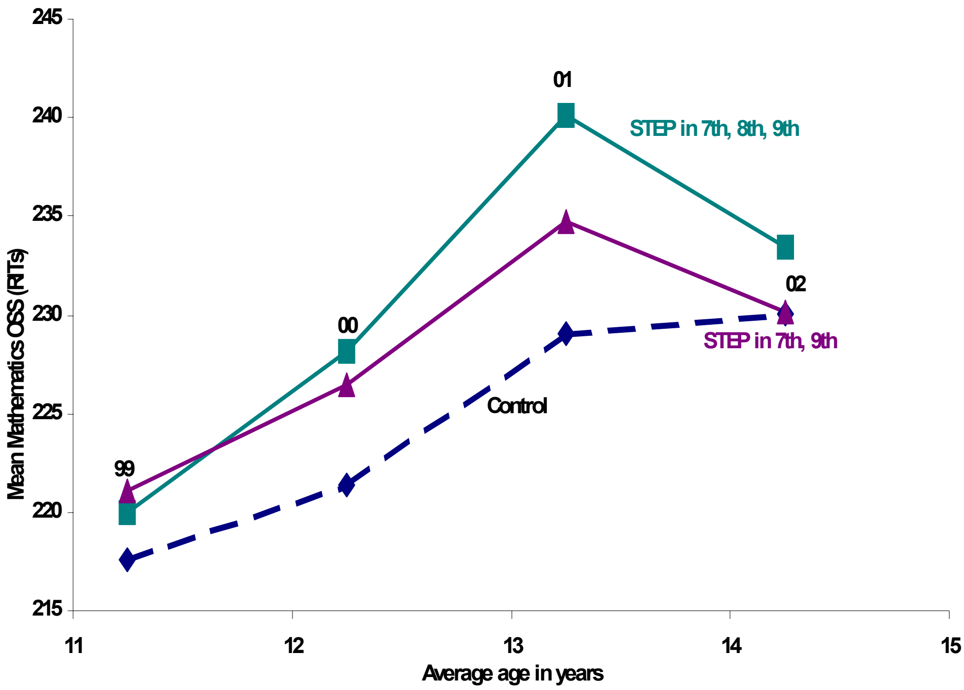


Figure 6.8 Mean Mathematics OSS of peers of Cohort B who missed STEP in 8th grade

important because it provides evidence that there are critical factors associated with the delivery of the CASE strategies that affect not only cognitive development but also Mathematics achievement.

Table 6.10
Mean Mathematics OSS of Cohort B students who missed STEP in 8th grade

		Mean Mathematics OSS in RITs (N students)		
Grade	**Average age (yr;mth)**	**Control**	**STEP in 6th, 7th & 8th grade**	**STEP in 7th & 9th grade only**
6th	**11; 3**	217.59 (198)	220.00 (84)	221.05 (22)
7th	**12; 3**	220.42 (188)	228.20 (97)	226.48 (25)
8th	**13; 3**	229.02 (223)	240.12 (100)	234.70 (23)
9th	**14; 3**	230.01 (185)	233.39 (101)	230.10 (22)

The second subset of students that experienced only part of the STEP intervention were peers of Cohort B students whose 8th grade teachers opted out of the STEP in the 2000-2001 school year (Table 6.10). In a one-sample, one-tailed t-test, no significant difference was found for the first or second years of the intervention, when comparing the change in Mathematics OSS of the part-STEP subset with the control A one-sample t-test was performed to compare the mean change in Mathematics OSS of students in Cohort B, who experienced all three years of the STEP intervention, with that of the mean change for students who missed the STEP in 8th grade. No significant difference was found in the first year, when both groups were receiving the STEP intervention, whereas in the second year there was a significant difference (t[df], p: 3.08 [32], <0.01) when the groups received different treatment. The effect size was 1.09 standard deviations in favour of the students who experienced three years of the intervention. This might indicate that the STEP intervention had enhanced the Mathematics achievement of these STEP students.

Reading & Literature achievement

The mean Reading & Literature Oregon State Scores (OSS) of students in 5th to 10th grades, tested in May 1999 prior to the STEP intervention, were used to create a cross-sectional control profile of Reading & Literature achievement (Figure 6.9a). The pattern shown by the control profile is one of little change from mean age 10+ to 11+, followed by a growth phase from mean age 11+ to 12+. The control sample met the state benchmarks for Reading & Literature in 5th and 8th grades.

Comparisons were made between the control profile for Reading & Literature and the achievement of the three cohorts of STEP students tested in May 1999, 2000, 2001 and 2002 (Figures 6.9 b, c, d). Cohort A met the 5th and 8th grade benchmarks, and started and finished close to the control profile. Cohort B also started at a point close to the control, but then rose above the control, meeting the 8th grade benchmark by a wider margin than did the control sample. Cohort C met the 8th grade benchmark, but not the 10th, in Reading & Literature.

Table 6.11
Mean Reading & Literature OSS of control

Grade (N students)	Average age (Yr; mth)	Mean OSS (RITs)
5th (43)	10; 3	220.53
6th (198)	11; 3	220.35
7th (192)	12; 3	222.63
8th (209)	13; 3	230.17
9th (185)	14; 3	231.29
10th (175)	15; 3	237.03

Table 6.12
Mean Reading & Literature OSS of STEP students

Grade	Average age (Yr;mth)	Mean Reading & Literature OSS in RITs (Nstudents)			
		Control	**Cohort A**	**Cohort B**	**Cohort C**
5th	**10; 3**	220.53 (43)	221.80 (46)		
6th	**11; 3**	220.35 (198)	222.20 (50)	222.00 (84)	
7th	**12; 3**	222.63 (192)	225.80 (50)	227.90 (100)	226.30 (55)
8th	**13; 3**	230.17 (209)	232.76 (50)	236.10 (101)	233.80 (61)
9th	**14; 3**	231.29 (185)		234.19 (86)	233.60 (56)
10th	**15; 3**	237.03 (175)			236.00 (53)

Regression analyses were performed of the pre-intervention (05/1999) and post-intervention (05/2002) Oregon State Scores. The results indicated that significant improvement had taken place in all three cohorts of STEP students over the period of the intervention (t[df], p Cohort A: 7.53 [45], <0.001; Cohort B: 9.57 [68], <0.001; Cohort C: 9.79 [37], <0.001). The effect sizes were large (Cohort A: 2.24 SDs, Cohort B: 2.23 SDs, Cohort C: 3.22 SDs).

The control slopes for Reading & Literature OSS were calculated from the means of the control sample data (as for the Mathematics OSS analysis) and are displayed in Table 6.13. Slopes were also calculated for STEP students and these are shown in Table 6.14. The control slopes were used as parameters in t-tests to compare overall change for control and STEP students. The mean change in the control Reading & Literature OSS over each period between testing was used as a parameter in t-tests to compare the annual change of control and STEP students. One-tailed probabilities were used in the t-tests because there is an *a priori* expectation that the treatment has a positive effect.

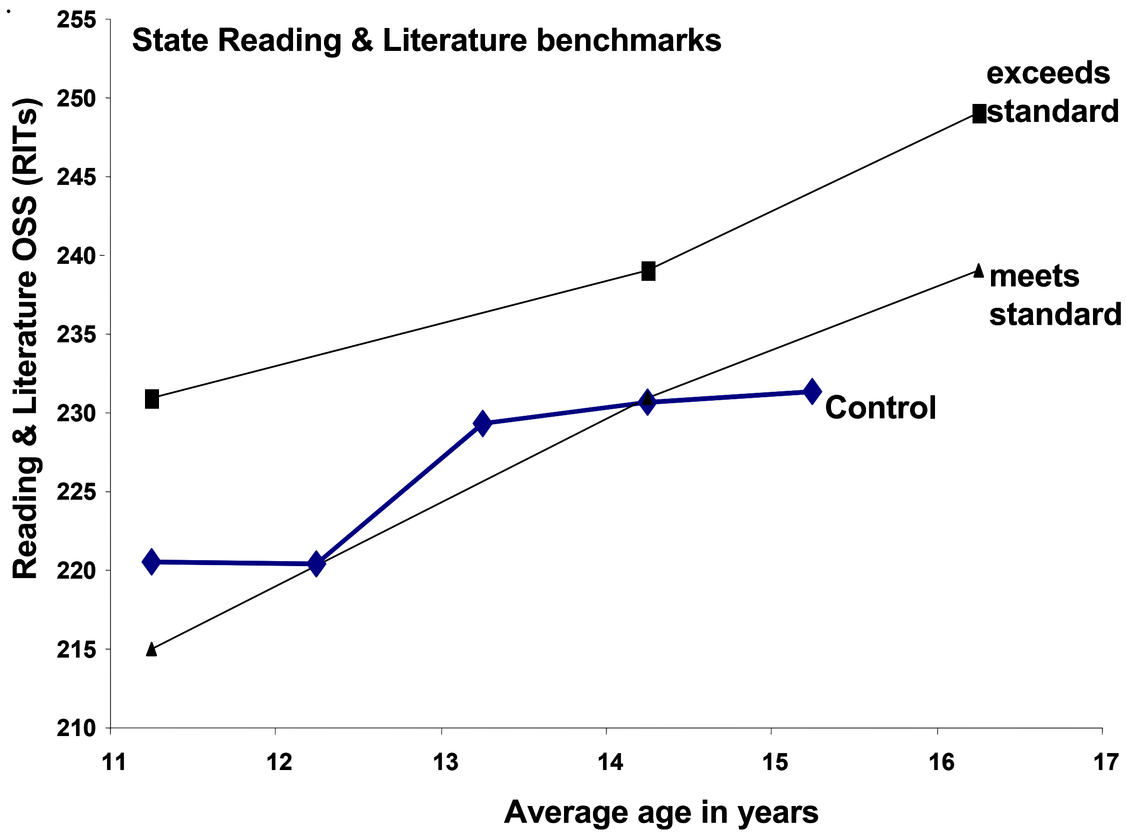


Figure 6. 9 a Mean Reading & Literature OSS of control

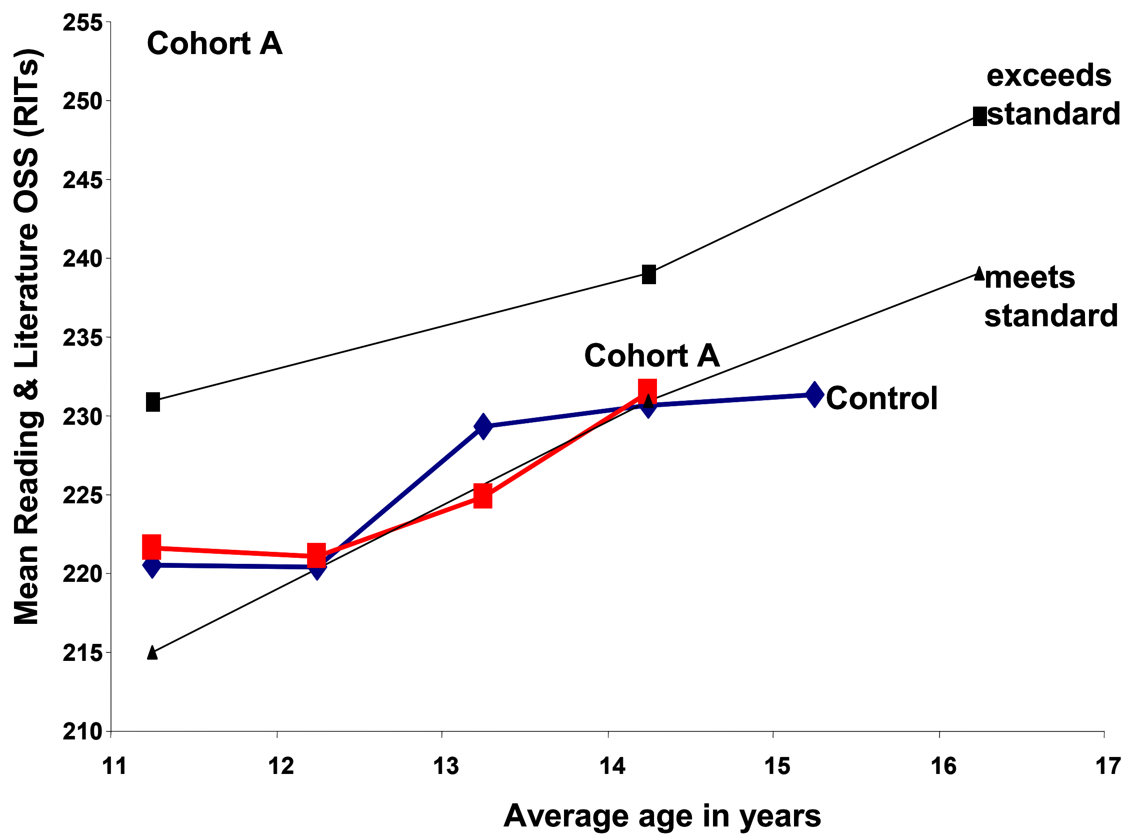


Figure 6. 9 b Mean Reading & Literature OSS of Cohort A

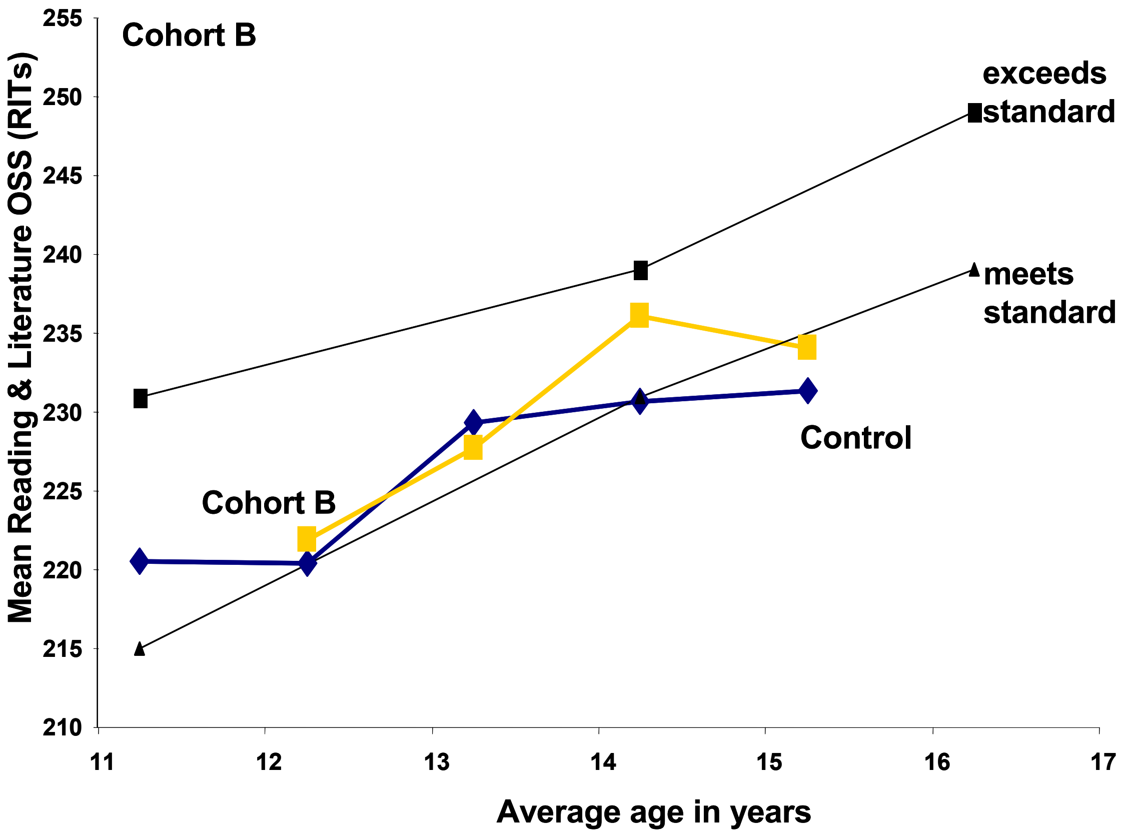


Figure 6. 9 c Mean Reading & Literature OSS of Cohort B

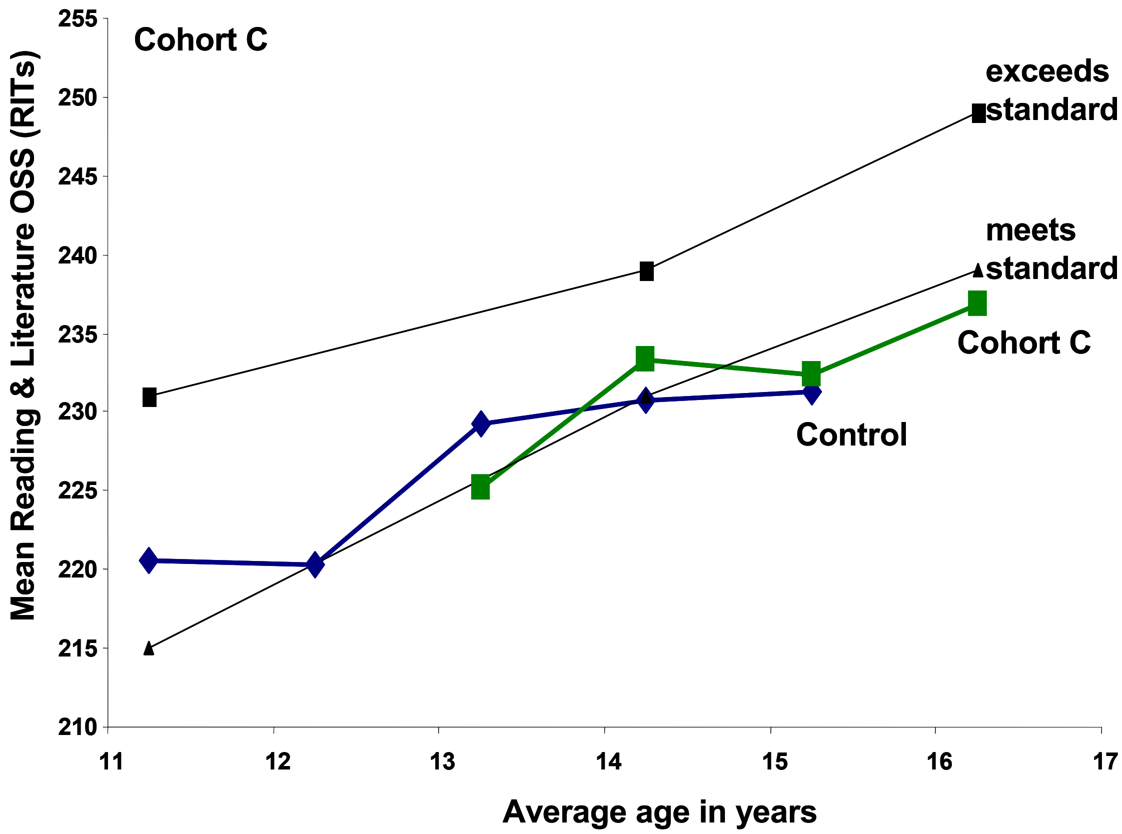


Figure 6. 9 d Mean Reading & Literature OSS of Cohort C

Table 6.13
Mean changes in Reading & Literature OSS of the control

	Mean in Reading & Literature OSS (RITs)			
Cohort equivalent	**Overall slope** (change in OSS/time in months)	**0-12 months**	**12-24 months**	**24-36 months**
A	0.27	-0.18	2.28	7.55
B	0.30	2.28	7.54	1.12
C	0.40	7.54	1.12	5.74

Table 6.14
Mean change in Reading & Literature OSS of STEP students (1999 to 2002)

	Mean change in Reading & Literature OSS in RITs (N students)			
Cohort	**Overall slope** (change in OSS/time in months)	**05/99 - 05/00**	**05/00 - 05/01**	**05/01 - 05/02**
A	0.31 (50)	0.65 (46)	3.56 (50)	6.96 (50)
B	0.35 (101)	5.27 (84)	8.25 (101)	2.71 (86)
C	0.29 (65)	7.46 (59)	0.61 (59)	5.18 (51)

There were no significant overall effects comparing the control and experimental slopes. However, Cohort B had a statistically significant large effect size gain (1.12 SDs) over the control in the first year of the STEP, followed by a statistically significantly loss (-1.19 SDs) in 2001- 2002.

A regression analysis was performed on the level of parent education and change in Reading & Literature OSS from 1999 to 2002, (t [N] = -0.70 [131], NS). No significant relationship was found, which suggests that the Reading & Literature achievement of these students is not associated with the educational level of their parents (Figure 6.10).

Table 6.15
Comparison of mean change in Reading & Literature OSS between control and STEP students

		Mean change in Reading & Literature OSS in RITs (N students)			
Cohort		**Overall 1999-2002**	**1999 - 2000**	**2000 - 2001**	**2001 - 2002**
A	*t*(df)	1.53 (49)	0.80 (45)	1.41 (49)	-0.50 (49)
	p	0.066 NS	0.21 NS	0.082 NS	0.31 NS
	Effect size *d*	0.44 (small)	1.12 (large)	0.40 (small)	-0.14 (nil)
B	*t*(df)	2.13 (100)	5.15 (83)	1.31 (100)	-5.55 (85)
	p	0.018 NS	<0.0001*	0.097 NS	<0.0001*
	Effect size *d*	0.43 (small)	1.12, large	0.26 (small)	-1.19, large
C	*t*(df)	-2.58 (64)	-0.18 (58)	-1.80 (58)	1.20 (50)
	p	0.0061 NS	0.43 NS	0.038 NS	0.12 NS
	Effect size *d*	-0.64 (medium)	0.05 (nil)	0.47 (small)	0.34 (small)

The *t*-tests are one-sample tests comparing control and experimental slopes (column1) and annual change (columns 2, 3, 4). ***t*** represents the *t* statistic, **df** represents degrees of freedom, and ***p*** represents probability of the null hypothesis. * indicates significance with the sequential Bonferroni test ($\propto = 0.05$).

When a regression analysis was performed on the Reading & Literature starting level (OSS 1999)and the change in Reading & Literature OSS from 1999 to 2002 (Figure 6.11), a significant relationship was revealed, suggesting that a lower starting level is associated with a greater gain (*t*[N]= -2.37 [121], *p* = <0.05). As with the corresponding Mathematics analysis, the regression of starting age and change in Reading & Literature OSS showed no relationship (*t*[N]= 0.29 [151], NS). There were individuals with high gains in Reading & Literature OSS across the whole age range (Figure 6.12).

The results for Reading & Literature achievement were analysed separately for male and female students. The mean Reading & Literature OSS for the males and females

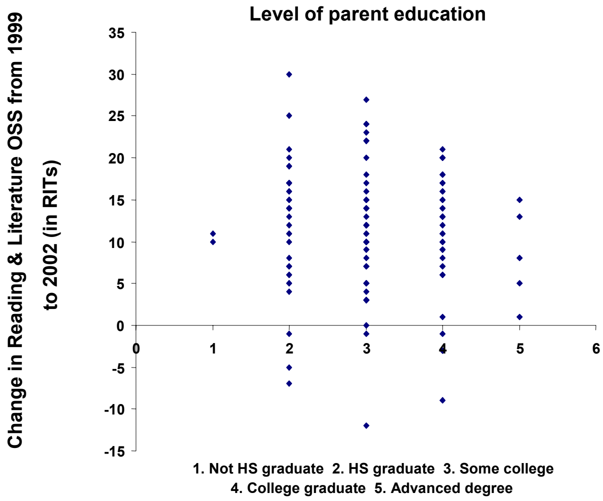


Figure 6.10 Level of parent education and change in Reading & Literature OSS

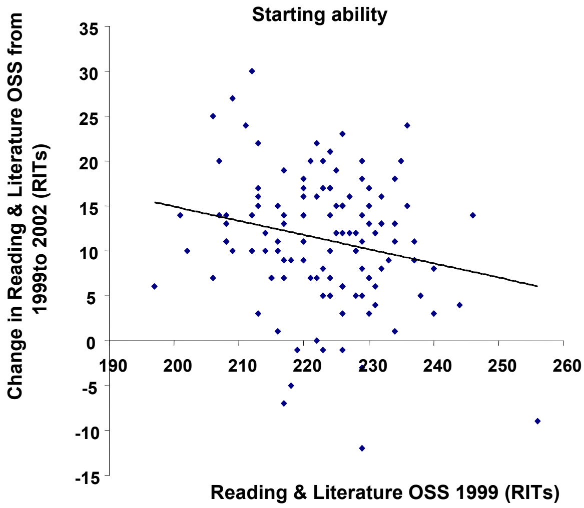


Figure 6.11 Starting ability and change in Reading & Literature OSS

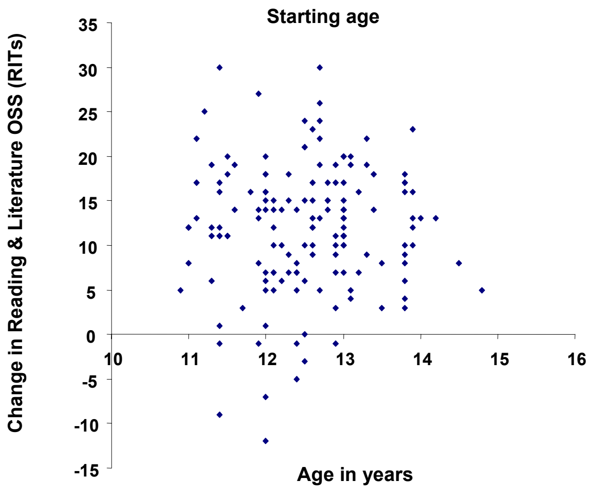


Figure 6.12 Starting age and change in Reading & Literature OSS

of each cohort are shown in Table 6.16, along with the means of the pre-intervention control population. Comparison of the male and female control students in the pre-intervention population revealed a similar pattern of language achievement for the sexes (Figure 6.13). There is little difference between males and females at mean age 10+, but after this point females slightly outperform their male peers. The growth spurt for both females and males occurs at mean age 12+ to 13+. The only significant difference between the Reading & Literature scores of males and females was in Cohort B in 1999, before the intervention began. Males performed better than females [*t*(df) 2.84 (57)], and the gain was of medium effect size (0.75 SDs).

Table 6.16
Mean Reading & Literature OSS of male and female students

			Mean Reading & Literature OSS in RITs (N students)			
Grade	**Average age (Yr;mth)**	**Sex**	**Control**	**Cohort A**	**Cohort B**	**Cohort C**
5th	**10; 3**	m	220.21 (24)	220.80 (24)		
		f	220.95 (19)	223.00 (22)		
6th	**11; 3**	m	220.00 (92)	221.10 (27)	225.10 (27) *	
		f	221.00 (109)	223.60 (23)	220.50 (57)	
7th	**12; 3**	m	221.00 (96)	225.00 (27)	229.30 (34)	225.00 (27)
		f	223.00 (92)	226.80 (23)	227.20 (66)	227.60 (28)
8th	**13; 3**	m	228.34 (98)	230.63 (27)	236.00 (35)	232.40 (31)
		f	232.00 (105)	235.26 (23)	236.20 (66)	235.20(30)
9th	**14; 3**	m	230.08 (94)		235.74 (27)	231.50 (30)
		f	232.00 (91)		233.47 (59)	236.00 (26)
10th	**15; 3**	m	235.24 (88)			236.31 (29)
		f	238.85 (87)			239.33 (24)

* indicates a significant difference between males and females.

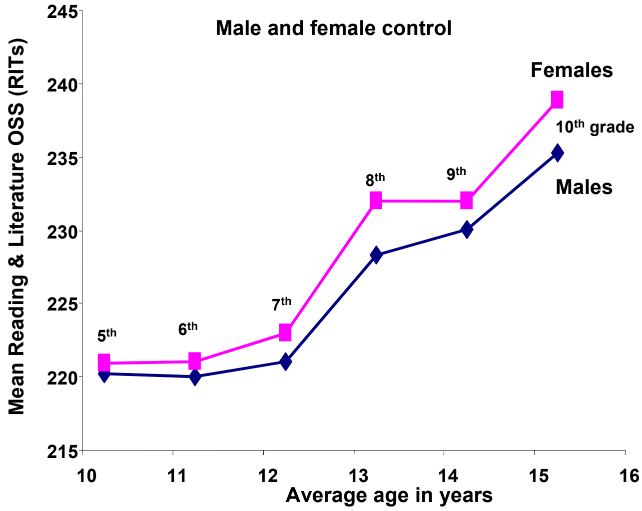
The results of one-tailed *t*-tests comparing the slopes of female and male STEP and control students are displayed respectively in Tables 6.17 and 6.18. One-tailed probabilities were used in the *t*- tests for males and females because of an *a priori* expectation that the treatment has a positive effect (Adey & Shayer, 1990, 1994). Overall, the slopes for Cohort B females showed a statistically significant gain (0.47 SDs) over the control, as well as a gain (1.42 SDs) in the first year of the intervention, followed by a significant loss (-0.83 SDs) in the third year. Cohort C females showed a statistically significant loss (-1.58 SDs) overall.

Table 6.17
Comparison of mean change in Reading & Literature OSS between female control and STEP students

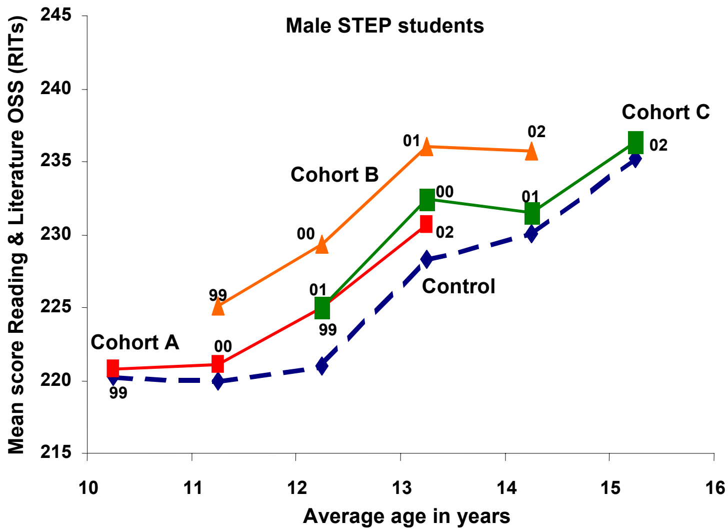
		Mean change in Reading & Literature OSS of females in RITs (N students)			
Cohort		**Overall (99-02)**	**05/99-05/00**	**05/00-05/01**	**05/01 -05/02**
A	***t*(df)**	0.09 (22)	0.38 (21)	0.76 (22)	-0.69 (22)
	p	0.46 NS	0.35 NS	0.23 NS	0.25 NS
	Effect size *d*	0.04 (nil)	0.16 (nil)	0.33 (small)	-0.29 (small)
B	***t*(df)**	1.88 (65)	5.25 (55)	-1.42 (65)	-3.16 (58)
	p	0.032*	<0.0001*	0.080 NS	0.0013*
	Effect size *d*	0.47, small	1.42, large	-0.35 (small)	-0.83, large
C	***t*(df)**	-4.39 (31)	-2.08 (29)	-0.09 (27)	1.24 (22)
	p	<0.0001*	0.023 NS	0.46 NS	0.11 NS
	Effect size *d*	-1.58, large	-0.77 (medium)	0.03 (nil)	0.53 (medium)

The *t*-tests are one-sample tests comparing control and experimental slopes (column1) and annual change (columns 2, 3, 4). *t* represents the *t* statistic, df represents degrees of freedom, and *p* represents probability of the null hypothesis. * indicates significance with the sequential Bonferroni test (∝ = 0.05).

a)



b)



c)

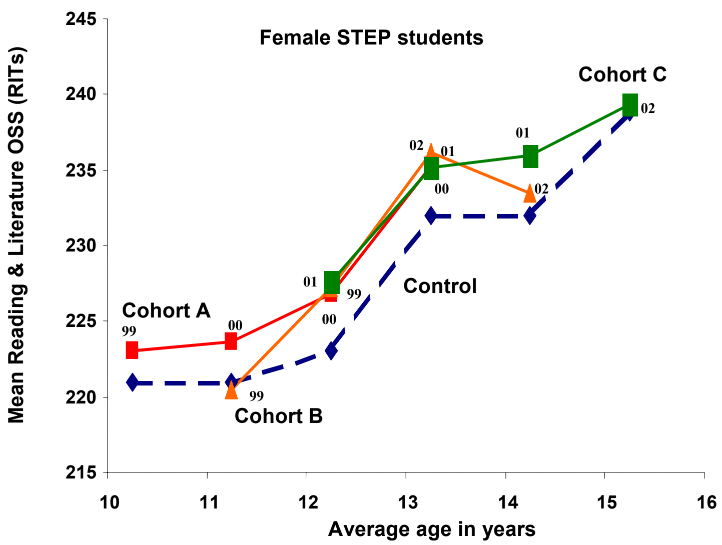


Figure 6.13 Mean Reading & Literature OSS of males and females

Males made no overall statistically significant gains over the control in Reading & Literature. However, males in Cohort A showed a large effect size gain (1.08 SDs) in the second year. Males in Cohort B had a significant gain (1.28 Sds) in 1999-2000, and a loss (-1.30 SDs) in the last year of the intervention.

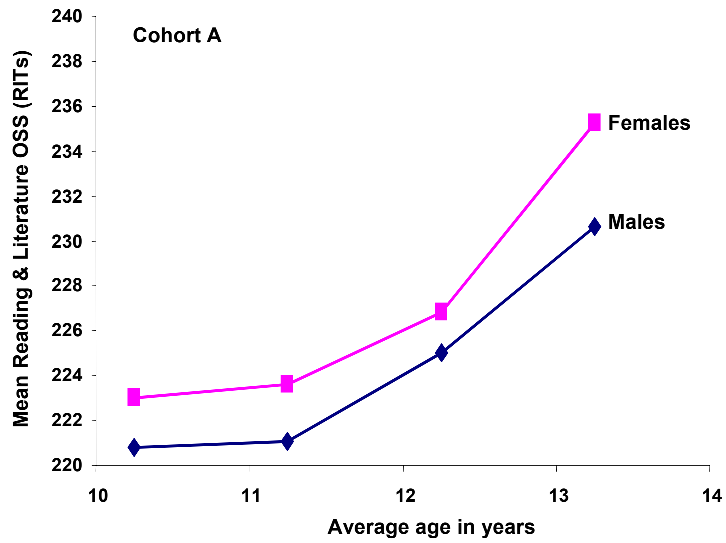
Table 6.18
Comparison of mean change in Reading & Literature OSS between male control and STEP students

		Mean change in Reading & Literature OSS of males in RITs (N students)			
Cohort		**Overall (99-02)**	**05/99 - 05/00**	**05/00 - 05/01**	**05/01 - 05/02**
A	***t*(df)**	1.64 (26)	0.60 (23)	2.74(26)	-1.23 (26)
	p	0.056 NS	0.28 NS	0.0055*	0.12 NS
	Effect size *d*	0.64 (medium)	0.25 (small)	1.08, large	-0.48 (small)
B	***t*(df)**	0.87 (34)	3.33 (27)	-0.38 (34)	-3.32 (26)
	p	0.19 NS	0.0013*	0.25 NS	0.0013*
	Effect size *d*	0.30 (small)	1.28, large	-0.13 (nil)	-1.30, large
C	***t*(df)**	-3.30 (32)	0.70 (28)	-1.89 (30)	0.38 (27)
	p	0.0012 NS	0.24 NS	0.034 NS	0.25 NS
	Effect size *d*	-1.17 (large)	0.26 (small)	-0.69 (medium)	0.14 (nil)

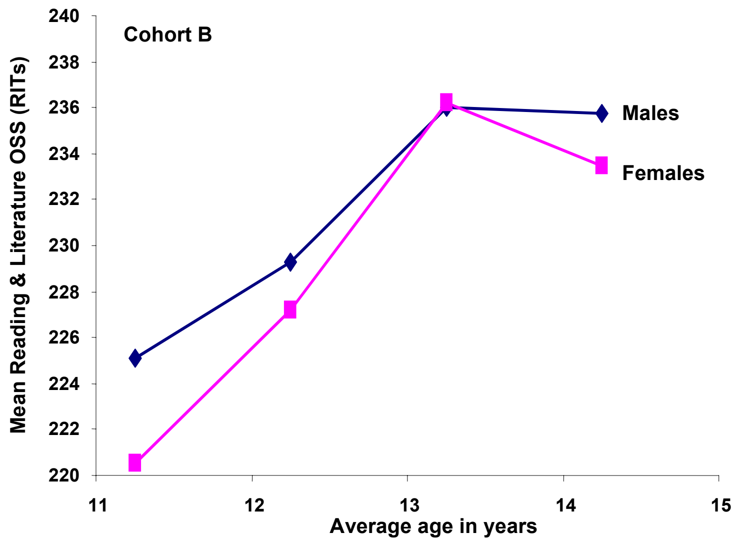
The *t*-tests are one-sample tests comparing control and experimental slopes (column1) and annual change (columns 2, 3, 4). ***t*** represents the *t* statistic, **df** represents degrees of freedom, and ***p*** represents probability of the null hypothesis. * indicates significance with the sequential Bonferroni test (∝ = 0.05).

A comparison of the Reading & Literature achievement of males and females for each cohort is shown in Figure 6.14. The pattern of Reading & Literature achievement of males and females in Cohort A is shown to be very similar, although the mean score for

a)



b)



c)

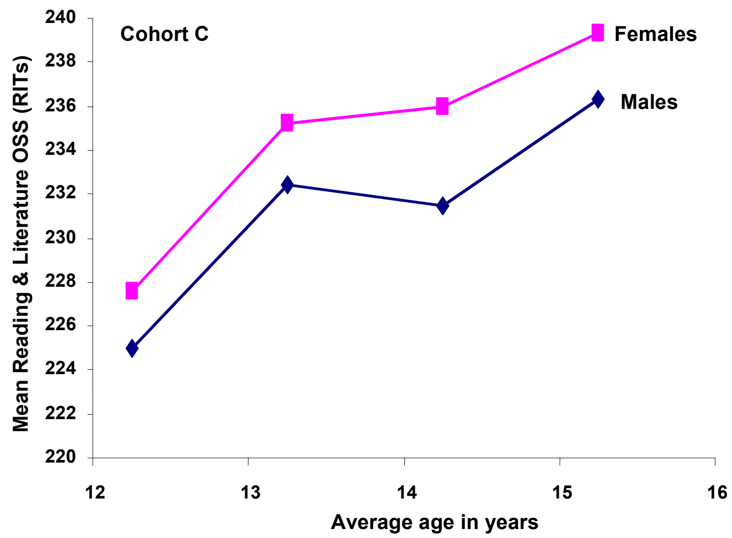


Figure 6.14 Mean Reading & Literature OSS of males and females of STEP cohorts

females is 3 to 4 RITs above those of their male peers. A similar trend to that of Cohort A is shown for males and females in Cohort C, where females also consistently outperform males. Cohort B shows a different trend, where males generally outperformed their female peers. However, there was no significant difference in annual change in Reading & Literature achievement between the males and females of any cohort.

The Reading & Literature results were analysed of those peers of Cohort A who did not experience the STEP intervention in 1999-2000 (see Table 6.19 and Figure 6.15). Some of these students, who began the STEP in 7th grade (September 2000), were new to the school, whereas others had been in classes that did not participate in STEP in 6th grade. The latecomers started out at a higher mean level than the students who had completed one year of the STEP intervention. Figure 6.15 suggests that the Cohort A closed the gap and almost caught up with the more able latecomers to the STEP by the end of 2002. However, when a one-tailed *t*-test was performed to compare the in the last two years of the intervention, no significant difference was found. Similarly, a two-tailed *t*-test of the difference between the means of the latecomers and their STEP peers also indicated that there was no significant difference between the means of the slopes from 2000 to 2002. This suggests that STEP intervention had little effect on the Reading & Literature achievement of these students.

The second subset of students that experienced only part of the STEP intervention were peers of Cohort B whose 8th grade teachers opted out of the STEP in the 2000-2001 school year (Table 6.20 and Figure 6.16). In a one-sample, one-tailed *t*-test, no significant difference was found for any year of the intervention, when comparing the change in Reading & Literature OSS of the part-STEP subset with the control. A two-sample *t*-test was performed to compare the mean change in Reading & Literature OSS of students in Cohort B, who experienced all three years of the STEP intervention, with that of the mean change for students who missed the STEP in 8th grade. No significant difference was apparent, which, again, suggests that the STEP intervention had little effect, if any, on the Reading & Literature achievement of these students.

Table 6.19
Mean Reading & Literature OSS of Cohort A students who began STEP in 7th grade

Grade	Average age (Yr; mth)	Mean Reading & Literature OSS in RITs (N students) Control	STEP in 6th, 7th & 8th	STEP in 7th & 8th
5th	**10; 3**	220.53 (43)	221.80 (46)	227.17 (18)
6th	**11; 3**	220.35 (198)	222.20 (50)	224.22 (18)
7th	**12; 3**	222.63 (192)	225.80 (50)	229.09 (22)
8th	**13; 3**	230.17 (209)	232.76 (50)	233.60 (20)

Table 6.20
Mean Reading & Literature OSS of Cohort B students who missed STEP in 8th grade

Grade	Average age (Yr; mth)	Mean Reading & Literature OSS (RITs)		
		Control	**STEP in 7th, 8th & 9th**	**STEP in 7th & 9th grade**
6th	**11; 3**	220.35 (198)	222.00 (84)	222.45 (22)
7th	**12; 3**	222.63 (192)	227.90 (100)	224.20 (25)
8th	**13; 3**	230.17 (209)	236.10 (101)	232.54 (22)
9th	**14; 3**	231.29 (185)	234.19 (86)	230.60 (22)

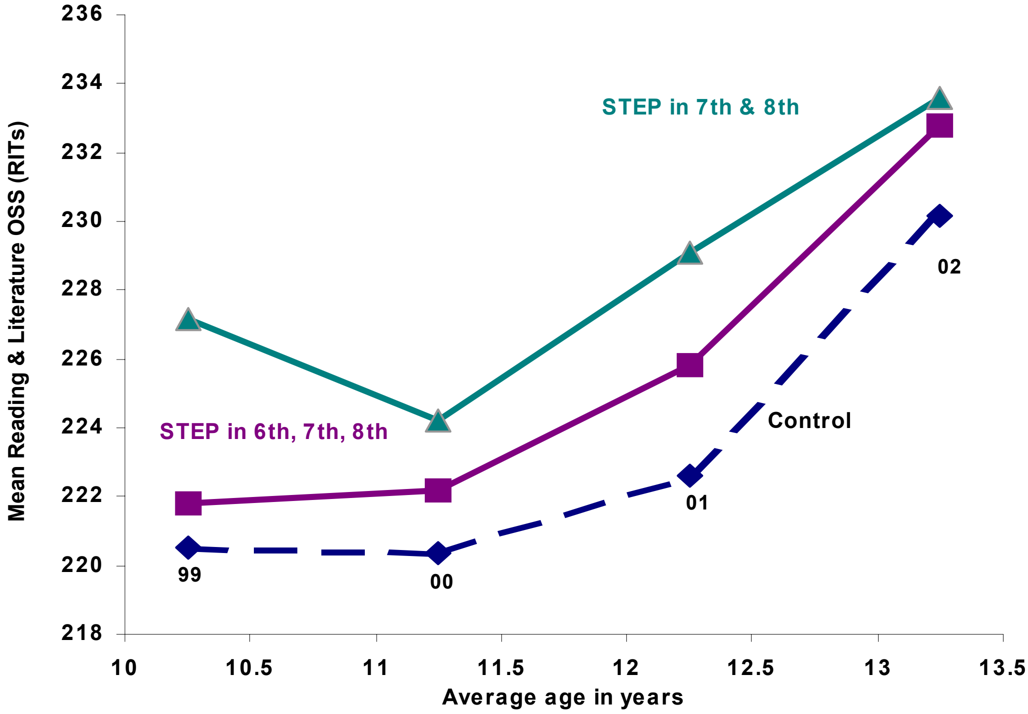


Figure 6.15 Mean Reading & Literature OSS of peers of Cohort A who missed STEP in 6th grade

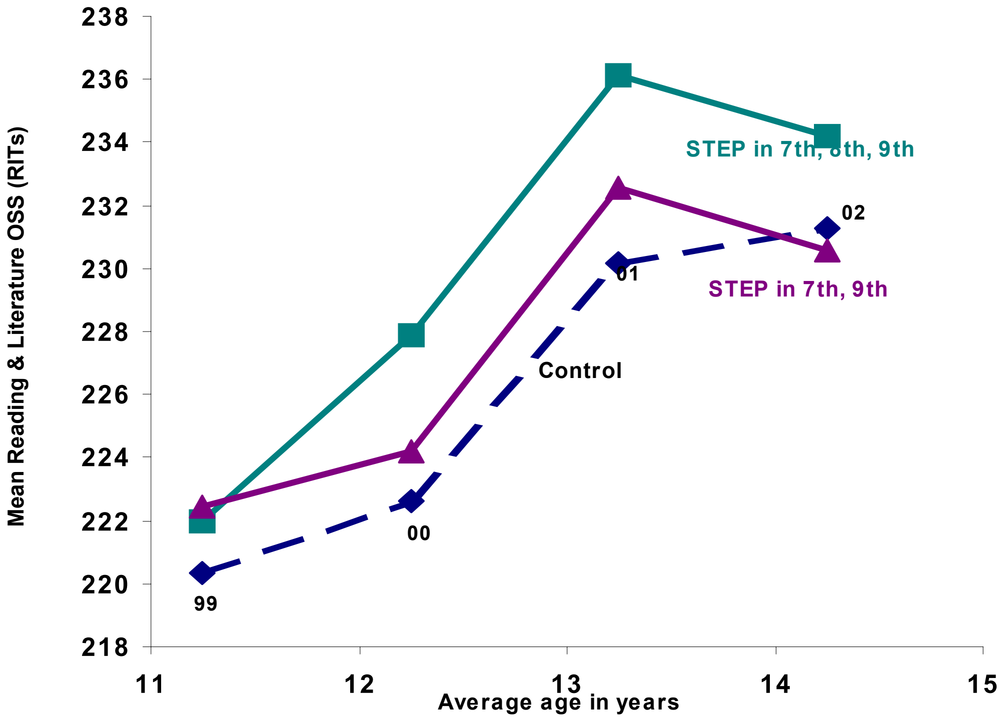


Figure 6.16 Mean Reading & Literature OSS of peers of Cohort B who missed STEP in 8th grade

Comparisons with students from Redwood and Sequoia School Districts

The Oregon State Test scores of the Molalla River students who experienced the STEP intervention were compared with those of students in two other school districts. To preserve anonymity, the comparison school districts have been given the pseudonyms Redwood and Sequoia. These districts were selected by the administrators of the Molalla River School District, because of their similar socioeconomic status to Molalla (see Figure 4.1, page 80). However, there were many unexplored variables between these schools,for example, school ethos, professional development emphasis, and curriculum offerings, that would also be likely to influence student achievement.

Redwood and Sequoia students were not exposed to CASE strategies during the period of this study. The agency that supplied the data, the Clackamas ESD, did not compile data from the Redwood and Sequoia middle schools prior to 2000, which meant that comparisons with Cohorts B and C were restricted to the period from 2000 to 2002.

Mathematics

The mean Mathematics OSS were calculated for the age-equivalent peers of the three cohorts of STEP students in the Redwood and Sequoia School Districts, and are shown in Table 6.21. Mean overall slopes and mean changes for each annual period were then calculated, and used as parameters in one sample *t*-tests to compare the overall slope and annual mean change in Mathematics OSS of STEP students with their age-equivalent peers in each of the other school districts.

No statistically significant overall gains were found for the STEP cohorts over two other school districts. Cohort A showed statistically significant losses against Redwood (-0.83 SDs) and Sequoia (-0.85 SDs) in 1999-2000, followed by a gain over Sequoia (0.93 SDs) in 2000-2001. Cohort B students made statistically significant, large effect size gains over both Redwood (1.15 SDs) and Sequoia (2.06 SDs) students from May 2000-May 2001, followed by significant losses against both schools (- 2.87 SDs and -1.98 SDs) in the following year. Cohort C showed a statistically significant loss (-0.96 SDs) against Sequoia in 2000-2001.

Table 6.21
Mean Mathematics OSS of STEP students and peers in Redwood and Sequoia School Districts

			Mean Mathematics OSS in RITs (N students)		
Cohort	**Grade**	**Average age (Yrs; mths)**	**STEP**	**Redwood**	**Sequoia**
A	5th	10; 3	221.40 (46)	224.05 (98)	222.43 (87)
	6th	11; 3	219.50 (50)	224.00 (111)	223.60 (94)
	7th	12; 3	226.24 (50)	231.00 (111)	227.30 (94)
	8th	13; 3	235.38 (50)	240.00 (106)	235.90 (85)
B	7th	12; 3	228.20 (97)	229.00 (116)	224.00 (110)
	8th	13; 3	240.12 (100)	237.00 (113)	230.00 (110)
	9th	14; 3	233.39 (101)	234.00 (105)	230.00 (103)
C	8th	13; 3	231.70 (61)	233.14 (113)	230.42 (93)
	9th	14; 3	230.12 (60)	231.58 (113)	231.12 (93)
	10th	15; 3	234.11 (45)	236.23 (99)	236.04 (87)

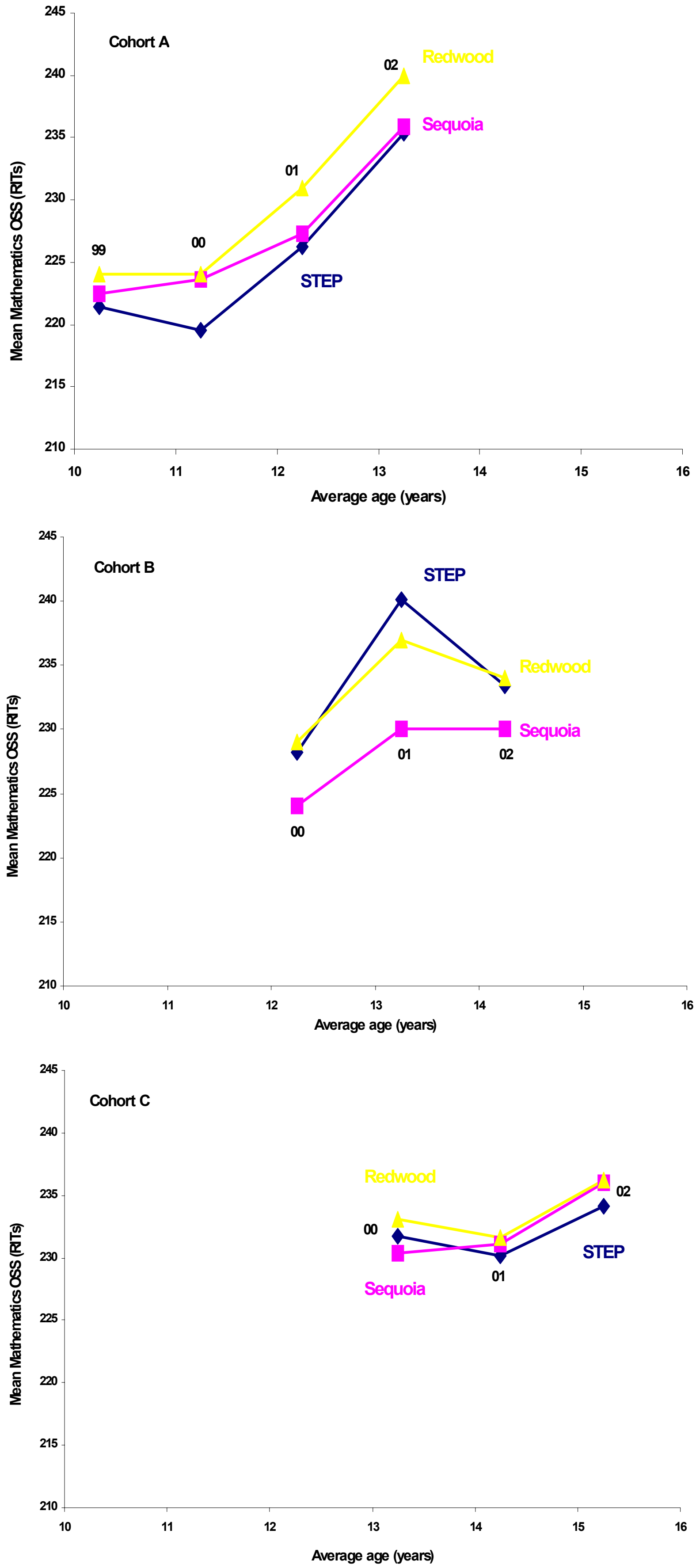


Figure 6.17 Mean Mathematics OSS of STEP cohorts and their Redwood and Sequoia peers

Table 6.22
Comparison of mean change in Mathematics OSS between STEP students and peers in Redwood and Sequoia School Districts

		Mean change in Mathematics OSS in RITs (N students)			
Cohort		**Overall 99-02**	**1999-2000**	**2000-2001**	**2001-2002**
A	t(df)	1.96 (49)	-2.86 (45)	3.27 (49)	1.18 (49)
STEP/ Sequoia	p	0.028, NS	<0.003*	<0.001*	0.12, NS
	Effect Size d	0.60 (medium)	-0.85 (large)	0.93, large	0.34 (small)
STEP/ Redwood	t(df)	-0.76 (49)	-2.78 (45)	-0.09 (49)	1.58 (49)
	p	0.23, NS	0.004*	0.46, NS	0.60, NS
	Effect Size d	-0.22 (small)	-0.83 (large)	- 0.03 (nil)	0.45 (small)
B	t(df)	-0.19 (97)	N/A	10.09 (96)	-9.83 (99)
STEP/ Sequoia	p	0.42, NS		<0.0001*	<0.0001*
	Effect Size d	-0.04 (nil)		2.06, large	-1.98, large
STEP/ Redwood	t(df)	0.34 (97)	N/A	5.63 (96)	-14.30 (99)
	p	0.37, NS		<0.0001*	<0.0001*
	Effect Size d	0.07 (nil)		1.15, large	-2.87, large
C	t(df)	-0.27 (51)	N/A	-3.68 (59)	0.95 (42)
STEP/ Sequoia	p	0.39, NS		0.0001*	0.17, NS
	Effect Size d	-0.08 (nil)		-0.96 (large)	0.29 (small)
STEP/ Redwood	t(df)	-0.50 (51)	N/A	-1.65 (59)	0.96 (42)
	p	0.31, NS		0.052 NS	0.17 NS
	Effect Size d	-0.14 (nil)		-0.43 (small)	0.23 (small)

The *t*-tests are one-sample tests comparing control and experimental slopes (column1) and annual change (columns 2, 3, 4). **N/A** indicates that no data are available. ***t*** represents the *t* statistic, **df** represents degrees of freedom, and ***p*** represents probability of the null hypothesis. ***** indicates significance with the sequential Bonferroni test ($\propto$ = 0.05).

Conclusions based on comparison with the non-project schools must necessarily be regarded as tentative. The Sequoia and Redwood School Districts had similar demographics to Molalla, but there were many unexplored variables between the schools that could also have influenced student achievement.

Reading & Literature

Corresponding data for the Reading & Literature achievement of the students in the three schools are displayed in Tables 6.23, 6.24, and Figure 6.18. No significant overall gains were found for STEP students over their peers in either school district. Cohort B exhibited statistically significant losses from 2001-2002 against Redwood (-0.31 SDs) and Sequoia (-1.19 SDs).

Table 6.23
Mean Reading & Literature OSS of STEP students and peers in Redwood and Sequoia School Districts

Cohort	**Grade**	**Average age (Yr; mth)**	**Mean Reading & Literature OSS (RITs)**		
			STEP	**Redwood**	**Sequoia**
A	**5th**	**10; 3**	221.80 (46)	223.00 (96)	222.00 (85)
	6th	**11; 3**	222.20 (50)	224.00 (111)	223.00 (85)
	7th	**12; 3**	225.80 (50)	228.00 (111)	226.00 (94)
	8th	**13; 3**	232.76 (50)	236.00 (106)	235.00 (97)
B	**7th**	**12; 3**	227.90 (100)	227.00 (111)	226.00 (118)
	8th	**13; 3**	236.10 (101)	234.00 (111)	230.00 (118)
	9th	**14; 3**	234.19 (86)	233.00 (100)	232.00 (107)
C	**8th**	**13; 3**	233.80 (61)	231.23 (111)	232.50 (96)
	9th	**14; 3**	233.60 (56)	231.67 (111)	231.94 (96)
	10th	**15; 3**	236.00 (53)	235.92 (101)	237.79 (85)

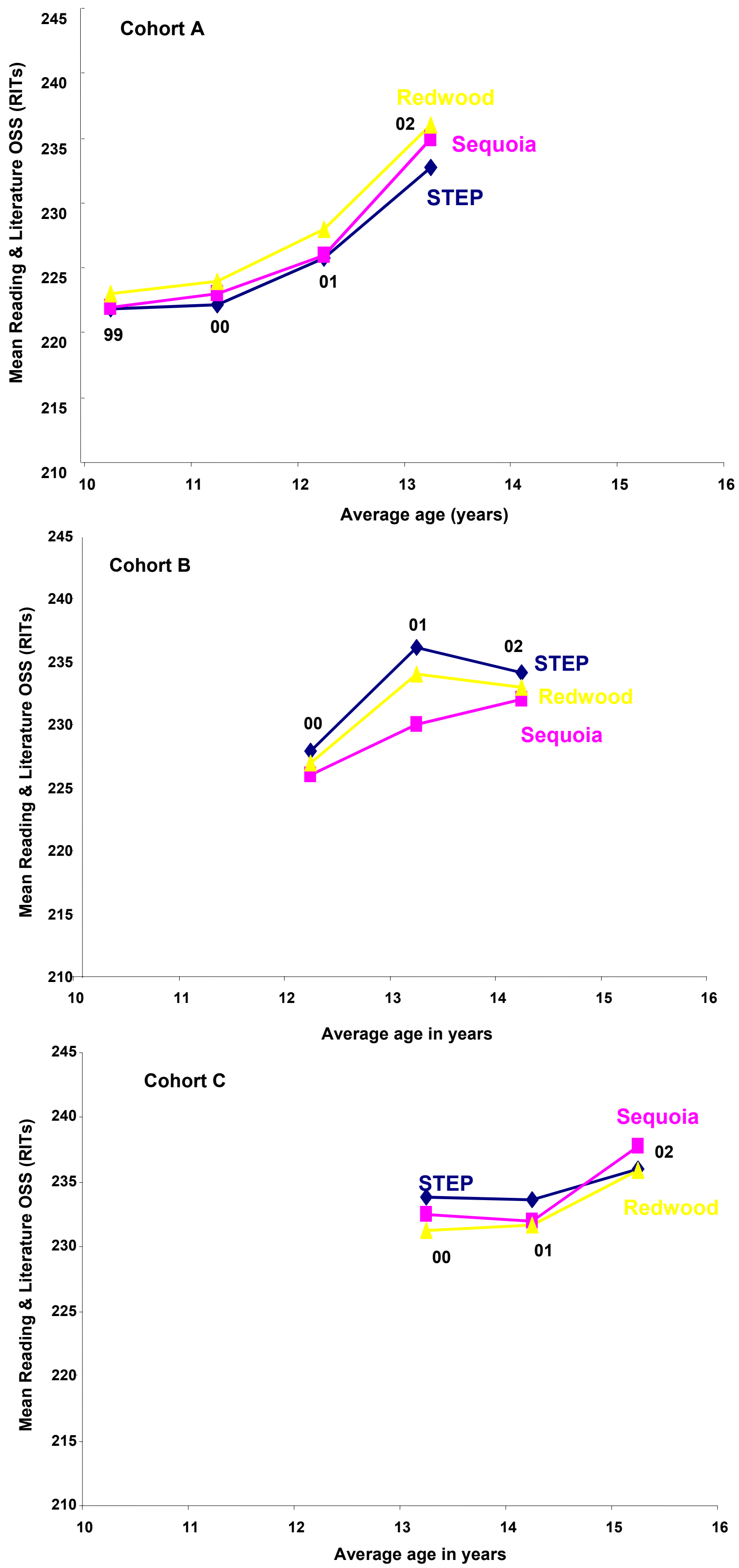


Figure 6.18 Mean Reading & Literature OSS of STEP cohorts and their Redwood and Sequoia peers

Table 6.24

Comparison of mean change in Reading & Literature OSS between STEP students and peers in Redwood and Sequoia School Districts

		Mean change in Reading & Literature OSS in RITs (N students)			
Cohort		**Overall (99-02)**	**05/99-05/00**	**05/00-05/01**	**05/01-05/02**
A	*t*(df)	0.71 (49)	0.15 (45)	0.94 (49)	-0.90 (49)
STEP/ Sequoia	*p*	0.24, NS	0.44, NS	0.18, NS	0.19, NS
	Effect size *d*	0.20 (small)	0.04 (nil)	0.27 (small)	-0. 26 (small)
STEP/ Redwood	*t*(df)	-1.70 (49)	-1.62 (45)	0.25 (49)	-0.65 (49)
	p	0.048 NS	0.056, NS	0.40, NS	0.26, NS
	Effect size *d*	-0. 49 (small)	-0. 48 (small)	0.07 (nil)	-0. 19 (nil)
B	*t*(df)	-0.72 (93)	N/A	1.75 (100)	-5.49 (85)
STEP/ Sequoia	*p*	0.24, NS		0.042, NS	<0.0001*
	Effect size *d*	-0.15 (nil)		0.35 (small)	-1.19. (large)
STEP/ Redwood	*t*(df)	-0.35 (93)	N/A	1.53 (100)	-6.96 (85)
	p	0.36, NS		0. 065, NS	<0.0001*
	Effect size *d*	-0.07 (nil)		0. (nil)	-0. 31 (small)
C	*t*(df)	-0.03 (54)	N/A	-1.34 (58)	-0.41 (50)
STEP/ Sequoia	*p*	0.5, NS		0. 093, NS	0.34, NS
	Effect size *d*	-0.01 (nil)		-0.35 (small)	-0.12 (nil)
STEP/ Redwood	*t*(df)	-0.31 (54)	N/A	-1.19 (58)	1.16 (50)
	p	0.5, NS		0.12, NS	0.13, NS
	Effect size *d*	-0.08 (nil)		-0.31(small)	0.33 (small)

The *t*-tests are one-sample tests comparing control and experimental slopes (column1) and annual change (columns 2, 3, 4). **N/A** indicates that no data are available. ***t*** represents the *t* statistic, **df** represents degrees of freedom, and ***p*** represents probability of the null hypothesis. ***** indicates significance with the sequential Bonferroni test ($\propto = 0.05$).

Science

A cross-sectional control profile for science was not plotted, because it was not possible to plot comparison profiles for the STEP cohorts. Cohorts A and B were tested on only one occasion, and Cohort C on two occasions during the period of the study, because statewide testing in Science takes place at 5th, 8th and 10th grades only.

The results of the 8th grade statewide Science testing, that occurred in 2000, 2001 and 2002 for STEP, Redwood and Sequoia students, are displayed in Table 6.25. The 8th grade benchmark for meeting the Science standard is 233 RITs, and for exceeding the standard 247 RITs. All of the results shown in Table 6.25 meet the benchmark standard, and none exceed it. It is interesting to note that the highest scores are those of Cohort B in 2001 (highlighted in Table 6.25), when these students had completed two school years of the STEP intervention.

Table 6.25
Mean Oregon State Scores for Science (8th grade)

	Mean Oregon State Scores for Science in RITs (N students)								
	2000			**2001**			**2002**		
	All	**Male**	**Female**	**All**	**Male**	**Female**	**All**	**Male**	**Female**
STEP Cohorts	C: 234.4 (81)	C: 234.5 (42)	C: 234.3 (39)	**B: 240.4 (75)**	**B: 244.5 (26)**	**B: 238.3 (49)**	A: 236.0 (36)	A: 235.0 (18)	A: 237.7 (18)
Redwood	233.1 (135)	233.0 (65)	233.3 (70)	237.3 (136)	237.6 (83)	236.8 (53)	238.0 (104)	238.0 (52)	237.6 (52)
Sequoia	237.0 (116)	239.2 (61)	234.6 (55)	235.0 (133)	236.8 (64)	233.3 (69)	237.0 (84)	240.0 (48)	234.1 (36)

Cohort C was the only group tested for Science on two occasions during the intervention. When the change in Science OSS from 2000 to 2002 was compared between STEP students and each of the other two schools, it was revealed that Cohort C performed significantly worse than both Sequoia and Redwood students [t(df) p: STEP/Sequoia, -2.34 (136), <0.05; STEP/Redwood, -2.16 (134), <0.05].

Relationship between cognitive development and school achievement

Correlations were calculated between the Oregon State Scores and the BLOT ability estimates for control students and for students who experienced the STEP intervention. The results of these correlation tests are shown in Table 6.26. Of the 84 correlations performed, 77 correlations were found to be significant at $p < 0.01$. The exceptions had sample sizes of 13 or less students. No statistically significant differences were found between the correlation coefficients of the control and any STEP cohort when tests for heterogeneity among two or more correlation coefficients were applied to the data in Table 6.26 (Sokal & Rohlf, 1995). The consistently strong correlation between the BLOT cognitive development estimate and the Oregon State Scores in Science, Mathematics, and Reading & Literature could be interpreted as evidence that cognitive development is associated with school achievement for both the control and STEP students. If the STEP correlations had been higher than those of the control, then one interpretation might be that STEP had enhanced the global cognitive functioning of the Molalla students. The CASE literature reports a gap of several years between the enhancement of cognitive development and significantly improved achievement scores (Adey & Shayer, 1990, 1991, 1994). Therefore, it is perhaps too early yet to find evidence of any such far-transfer effects of the STEP in the achievement results of the STEP students.

Table 6.26

Correlation between BLOT ability estimate and Oregon State Scores

			Correlation Coefficient *r* (N), *p*		
Tests			**Cohort A**	**Cohort B**	**Cohort C**
Mathematics OSS vs BLOT	Control	all	0.77 (22), <0.01	0.73 (71), <0.01	0.87 (21), <0.01
		males	0.74 (9), <0.05	0.67 (25), <0.01	0.95 (13), <0.01
		females	0.74 (13), <0.01	0.74 (46), <0.01	0.82 (8), <0.01
	2000	all	0.74 (43), <0.01	0.71 (131), <0.01	0.75 (63), <0.01
		males	0.69 (19), <0.01	0.67 (62), <0.01	0.80 (31), <0.01
		females	0.79 (24), <0.01	0.77 (69), <0.01	0.72 (32), <0.01
	2001	all	0.70 (36), <0.01	0.59 (66), <0.01	0.60 (74), <0.01
		males	0.70 (17), <0.01	0.64 (26), <0.01	0.66 (37), <0.01
		females	0.73 (19), <0.01	0.57 (40), <0.01	0.55 (37), <0.01
	2002	all	0.79 (35), <0.01	0.63 (93), <0.01	0.61 (28), <0.01
		males	0.78 (21), <0.01	0.52 (32), <0.01	0.80 (16), <0.01
		females	0.88 (14), <0.01	0.68 (61), <0.01	0.45 (12), NS
Reading & Literature OSS vs BLOT	Control	all	0.79 (21), <0.01	0.75 (71), <0.01	0.72 (21), <0.01
		males	0.78 (8), <0.05	0.67 (25), <0.01	0.67 (13), <0.05
		females	0.80 (13), <0.01	0.77 (46), <0.01	0.79 (8), <0.05
	2000	all	0.75 (43), <0.01	0.76 (143), <0.01	0.74 (63), <0.01
		males	0.81 (19), <0.01	0.77 (65), <0.01	0.73 (32), <0.01
		females	0.71 (24), <0.01	0.75 (78), <0.01	0.78 (31), <0.01
	2001	all	0.71 (36), <0.01	0.60 (65), <0.01	0.65 (71), <0.01
		males	0.75 (17), <0.01	0.63 (24), <0.01	0.70 (35), <0.01
		females	0.65 (19), <0.01	0.60 (41), <0.01	0.61 (36), <0.01
	2002	all	0.79 (35), <0.01	0.63 (86), <0.01	0.61 (34), <0.01
		males	0.80 (21), <0.01	0.64 (27), <0.01	0.66 (18), <0.01
		females	0.77 (14), <0.01	0.61 (59), <0.01	0.56 (16), <0.01
Science OSS vs BLOT	Control	all		0.75 (16), <0.01	
		males		0.75 (10), <0.05	
		females		0.79 (6), NS	
	2000	all			0.76 (42), <0.01
		males	N/A	N/A	0.73 (20), <0.01
		females			0.77 (22), <0.01
	2001	all	N/A	0.66 (94), <0.01	
		males		0.68 (33), <0.01	N/A
		females		0.68 (61), <0.01	
	2002	all	0.76 (35), <0.01		
		males	0.84 (21), <0.01	N/A	N/A
		females	0.66 (14), <0.01		

N/A: no data available (statewide testing of science in Oregon occurs only in 5th, 8th and 10th grades.)

CHAPTER 7

What new skills did STEP students acquire?

One advantage of applying Rasch analysis to the results of the BLOT tests is that the acquisition of new formal operational skills by the STEP students can be followed over time. Comparison of the item-person map of the 1999 BLOT data (Figure 5.1) with the content of the BLOT (Table 2.2) allows particular formal operational skills to be ascribed to developmental levels. The 1999 BLOT ability estimates, raw score equivalents, BLOT item numbers, and particular formal operations are displayed as a BLOT scale in Table 7.1. As the results of all subsequent BLOT tests were anchored to the item estimates of the data of 1999, the BLOT scale in Table 7.1 is correct for all student samples in this study.

The BLOT scale reveals that a change in the BLOT ability estimate of, say, 0.50 logits would indicate that particular formal operational skills have been acquired, depending on the actual starting and finishing developmental level of the individual. For example, if a student's score improved by 0.5 logits from 1.0 to 1.5, then it is highly probable that the student has added the operations complete affirmation, implication and negation of *q* to that student's formal operational repertoire. In contrast, a gain of 0.5 logits by those whose score increased from -1.0 to -0.5 would indicate that different formal operational skills had most likely been acquired, which, in this case, would be the more advanced operation of correlations, probability, and negation (to negate correlative).

Table 7.1
BLOT scale

BLOT ability estimate (logits)	BLOT raw score equivalent	BLOT item #	Operations
4.91	35	32	Negation of reciprocal implication
3.77	34	21	Correlative + negation → equilibrium
3.04	33	30	Equivalence
2.58	32	28	Non- implication
2.25	31	15	Reciprocal implication
1.98	30	13	Reciprocal exclusion
1.75	29	8	Correlations
1.54	28	17	Identity (to negate reciprocal)
1.36	27	31	Negation of *q*
1.19	26	3	Implication
1.03	25	26	Complete affirmation
0.88	24	35	Coordination of two systems of reference
0.74	23	9	Conjunction
0.60	22	19	Reciprocal (to cause disequilibrium)
0.46	21	29	Affirmation of *q*
0.33	20	25	Complete negation
0.20	19	23	Correlative + identity → disequilibrium
0.07	18	11	Conjunctive negation
-0.06	17	10	Disjunction
-0.19	16	34	Coordination of two systems of reference
-0.32	15	22	Reciprocal + negation → disequilibrium
-0.45	14	5	Multiplicative compensation
-0.59	13	20	Negation (to cause disequilibrium)
-0.73	12	18	Negation (to negate correlative)
-0.87	11	33	Probability
-1.02	10	7	Correlations
-1.19	9	4	Incompatibility
-1.36	8	16	Reciprocal (to negate identitiy)
-1.54	7	24	Coordination of two systems of reference
-1.75	6	14	Probability
-1.99	5	27	Negation of *p*
-2.26	4	1	Negation (to negate identity)
-2.60	3	2	Reciprocal (to negate identity)
-3.06	2	12	Affirmation of *p*
-4.00	1	6	Correlations

The mean BLOT ability estimate of Cohort A changed from -0.052 logits in 09/1999 to 1.432 in 05/2002. Table 7.1 reveals that the change in mean BLOT ability estimate of Cohort A most probably represents an acquisition of 11 new formal operational skills, on average, over the period of the study: disjunction, conjunctive negation, correlative + identity → disequilibrium, complete negation, affirmation of *q*, reciprocal (to cause disequilibrium), conjunction, coordination of two systems of reference, complete affirmation, implication, and negation of *q*. This acquisition, on average, of eleven new formal operational skills between average age 11 years six months and 14 years 3 months, is in accord with the notion that early adolescence is a time of rapid cognitive development for most children (Shayer & Adey, 1981).

During the same time interval, the students in Cohort B showed mean cognitive growth from 0.778 to 1.914 logits, indicating an average gain of six new skills to their formal operational repertoire. Between average age 12 years 6 months and 15 years 3 months, the students acquired coordination of two systems of reference, complete affirmation, implication, negation of *q*, identity (to negate reciprocal), and correlations.

The mean cognitive growth for Cohort C students (0.951 to 1.866 logits) represented an average gain of five operations: complete affirmation, implication, negation of *q*, identity (to negate reciprocal), and correlations. The cognitive growth, from average ages 13 years six months to 16 years 3 months for Cohort C, would not be predicted from the data of the representative British surveys, that found no increase in the proportion of students showing formal operational thinking beyond the age of 15 years (Shayer & Wylam, 1978). The late-onset cognitive growth found in Cohorts B and C might support the conclusion that enhancement of these students' cognitive development had occurred.

Individual vignettes

Vignettes of individual students are presented, in order to illustrate some of the more interesting trends in the data presented in Chapters 5 and 6. Pseudonyms have been used to preserve the anonymity of these students. Marty was selected as a student typical of Cohort A, because of the similarity of his profile of cognitive development to that of the control. Rose, also in Cohort A, was a more advanced peer of Marty, chosen to illustrate the remarkable heterogeneity in cognitive development. Vicky, a student in Cohort B, was selected because she showed the late-onset cognitive growth spurt common to many of her peers. Arthur (also Cohort B) was chosen because his BLOT results showed the fall during 8th grade characteristic of many males, as well as a fall in mathematics achievement in the following year. Ron, a student in Cohort C, was selected because his cognitive profile also displayed a fall during 8th grade. All of these students were Caucasian.

Marty

Marty was a student in Cohort A, aged 11 years 5 months in September 1999. The level of education of his parents was undeclared. Seventeen CASE activities were delivered to Marty's cohort in the course of the intervention. In common with many of his peers, Marty acquired many more new formal operational skills during 6th grade than in the following two school years (see Table 7.2). Figure 7.1a reveals that the pattern of Marty's cognitive development was similar to that of the control profile. Marty's cognitive development was estimated to be at the early concrete operational stage in September 1999, and he had attained the early formal operational level by the end of the period of study.

Table 7.2 *Cognitive development profile of Marty*

Change in BLOT ability estimate (logits)	Additional BLOT questions correctly answered	New formal operations acquired
i) 0.09 to 1.03	1, 7, 9, 12, 14, 16, 17, 20, 22, 25, 35	Negation (to negate identity), correlations, conjunction, affirmation of *p*, probability, reciprocal (to negate identity), identity (to negate reciprocal), negation (to cause disequilibrium), reciprocal+negation→disequilibrium, complete negation, and coordination of two systems of reference.
ii) 1.03 to 1.36	26, 30, 33, 34	Complete affirmation, equivalence, probability, and coordination of two systems of reference.
iii) 1.36 to 1.75	4, 8, 10, 15	Incompatibility, correlations, disjunction, and reciprocal implication.

From May 1999 to May 2002, Marty's Mathematics OSS increased by 22 RITs (1999:216, 2000:223, 2001:230, 2002:238). When Marty's Mathematics achievement is compared with the benchmarks for the Oregon Mathematics standards (Figure 7.1b), it is clear that Marty's achievement has improved from just meeting the standard at 5th grade to almost exceeding the standard by 8th grade. Marty was also closer to the benchmark for exceeding the standard in Reading & Literature OSS in May 2002 than in May 1999 (Figure 7.1c). His Reading & Literature OSS improved by 11 RITs (225, 220, 223, 236) over the three-year period. Marty's 8th grade Science OSS was 240, also indicative of meeting the state standard for that grade.

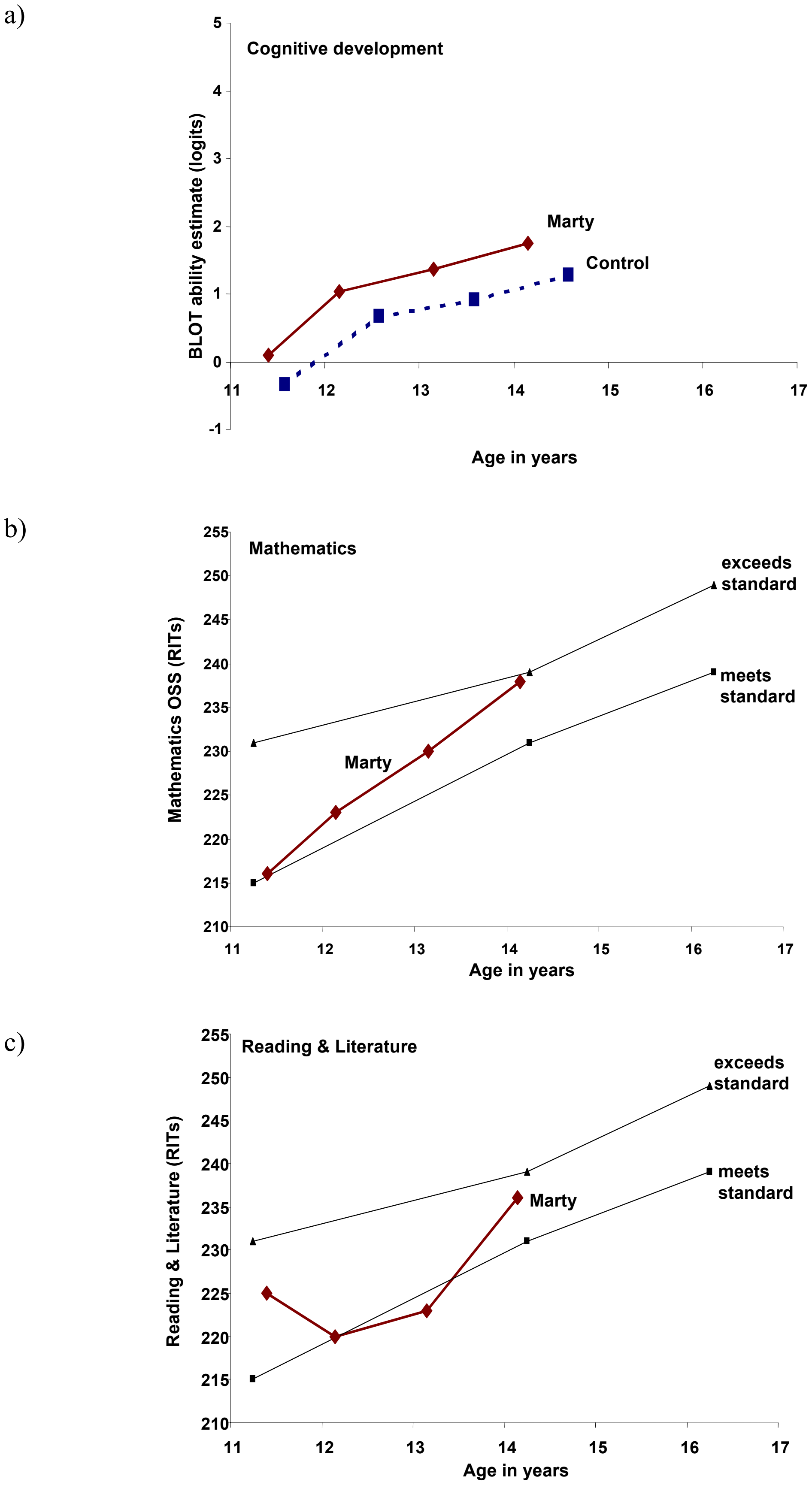


Figure 7.1 Cognitive development and achievement profiles for Marty

Rose

Rose was a peer of Marty in Cohort A, aged 11 years 7 months in September 1999. The level of education of her parents was reported as 3 (some college education). Rose's level of cognitive development prior to the intervention was estimated to be at the early formal operational stage, and she had reached the mature formal operational level by May 2002. Figure 7.2a indicates that Rose was one of the more cognitively developed children in Cohort A prior to the intervention and, accordingly, she had relatively fewer new operations to acquire compared with more typical students such as Marty.

Table 7.3 *Cognitive development profile of Rose*

Change in BLOT ability estimate (logits)	**Additional BLOT questions correctly answered**	**New formal operations acquired**
i) 2.00 to 1.54	8, 25	Correlations, complete negation.
ii) 1.54 to2.25	10, 13, 31, 34	Disjunction, reciprocal exclusion, negation of *q*, and coordination of two systems of reference.
iii) 2.25 to 3.77	21, 26	Correlative + negation → equilibrium, and complete affirmation.

Rose's Mathematics OSS increased by 13 RITs (233, 223, 235, 246) over the period of the study, and she exceeded the 5th and 8th grade standards in both Mathematics and Reading & Literature (Figure 7.2b). However, Rose did not show as strong an improvement in Reading & Literature (Figure 7.2c), and her 2002 score was 4 RITs less than in 1999 (240, 220, 223, 236). Her 8th grade Science OSS was 240, which met, but did not exceed, the Oregon standard.

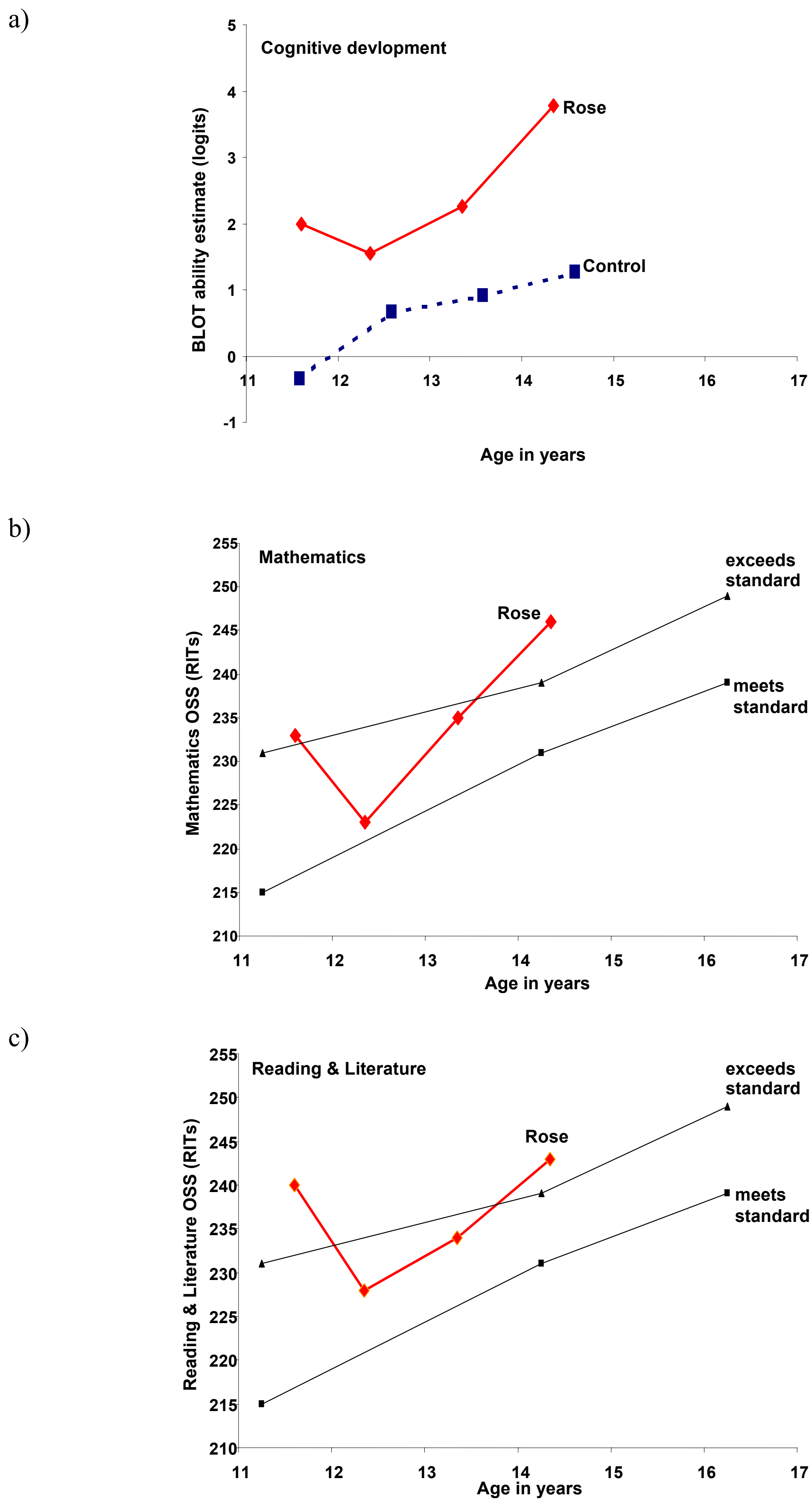


Figure 7.2 Cognitive development and achievement profiles for Rose

Vicky

Vicky was a student in Cohort B, whose BLOT results (Figure 7.3a) showed the late-onset cognitive spurt typical of many of her peers. Vicky was aged 12 years six months at the start of the STEP intervention and her parents' level of education was 2 (high school graduate). She was a member of the sub-group of Cohort B that experienced 16 CASE activities during the three-year period of the STEP. Vicky's cognitive development was estimated to be at the early concrete operational level at the start of the study, and had progressed to the early formal operational stage by May 2002.

Table 7.4 *Cognitive development profile of Vicky*

Change in BLOT ability estimate (logits)	Additional BLOT questions correctly answered	New formal operations acquired
i) -0.59 to 0.33	8, 9, 10, 25, 26, 28, 32	Correlations, conjunction, disjunction, complete negation, complete affirmation, non- implication, and negation of reciprocal implication.
ii) 0.33 to 0.74	6, 11, 17, 18, 20, 22, 23, 30	Correlations, conjunctive negation, identity (to negate reciprocal), negation (to negate correlative), negation (to cause disequilibrium), reciprocal + negation → disequilibrium, correlative + identity → disequilibrium, and equivalence.
iii) 0.74 to 1.75	7, 15, 19, 29, 31, 35	Correlations, reciprocal implication, reciprocal (to cause disequilibrium), affirmation of *q*, negation of *q*, and coordination of two systems of reference.

Vicky had not yet met the state standard for Mathematics during the period of the study (Figure 7.3b), although her OSS increased by 11 RITs (215, 218, 227, 226). She made more substantial gains (21 RITS) in Reading & Literature (224, 226, 242,

a)

b)

c)

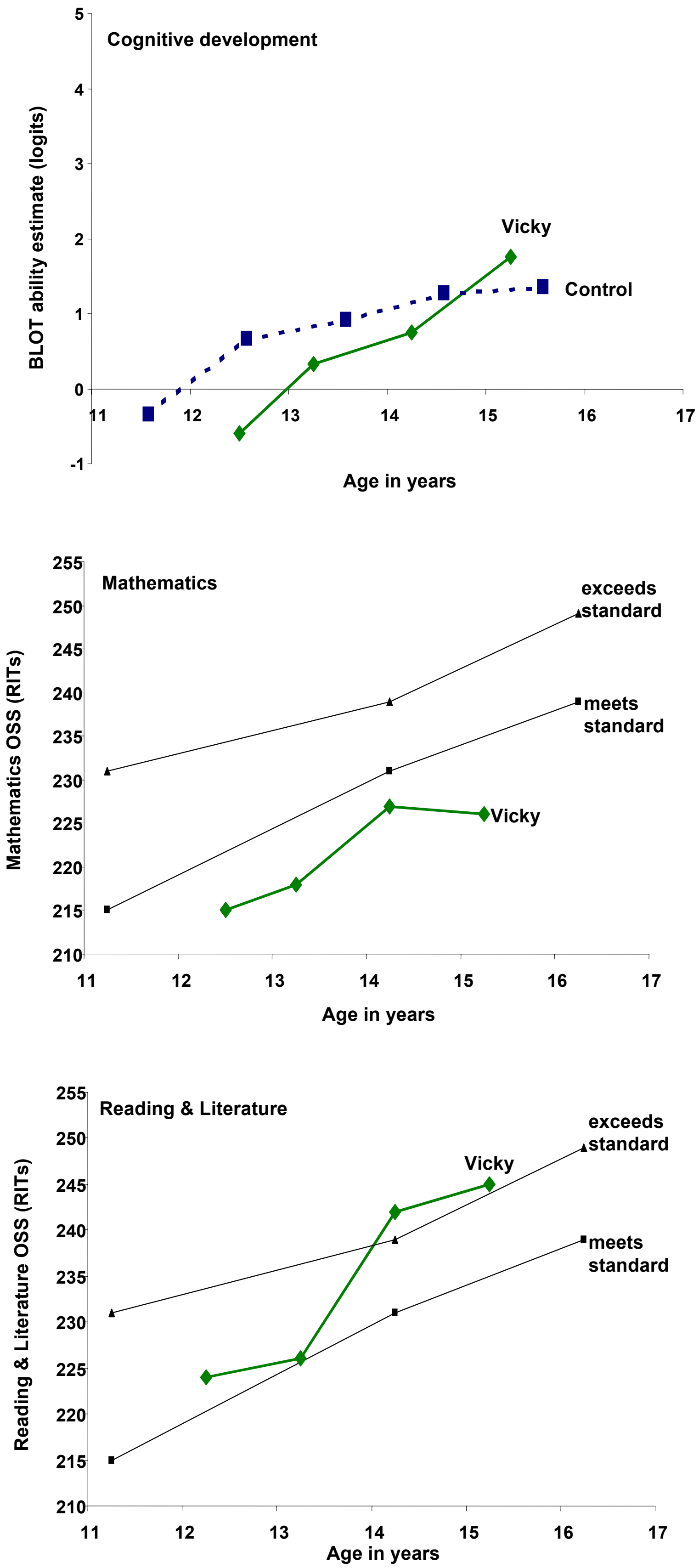


Figure 7.3 Cognitive development and achievement profiles for Vicky

245). Vicky exceeded the 8th grade state standard in both Reading & Literature and Science (254).

Arthur

Arthur's results showed the fall in BLOT ability estimate from 2000 to 2001 that was characteristic of many of his male peers in Cohort B. He was aged 12 years 5 months at the start of the STEP intervention. The level of his parents' education was 5 (advanced degree). Arthur was a member of the sub-group of Cohort B that experienced 21 CASE activities during the three-year period of the STEP intervention, and his cognitive development profile is displayed in Table 7.4. His cognitive developmental level was estimated to be at the mature concrete stage in September 1999, and he had progressed to the mature formal operational level by May 2002. Figure 7.4a reveals that Arthur showed a remarkable late cognitive growth spurt at age 14+.

Table 7.5 *Cognitive development of Arthur*

Change in BLOT ability estimate (logits)	Additional BLOT questions correctly answered	New formal operations acquired
i) 1.38 to 3.04	3, 14, 20, 21, 22, 23, 25	Implication, probability, negation (to cause disequilibrium), correlative + negation → equilibrium, correlative + identity → disequilibrium, complete negation.
ii) 3.04 to -0.20	no additional correct answers	N/A
iii) -0.20 to 4.84	13, 26	Reciprocal exclusion, complete affirmation.

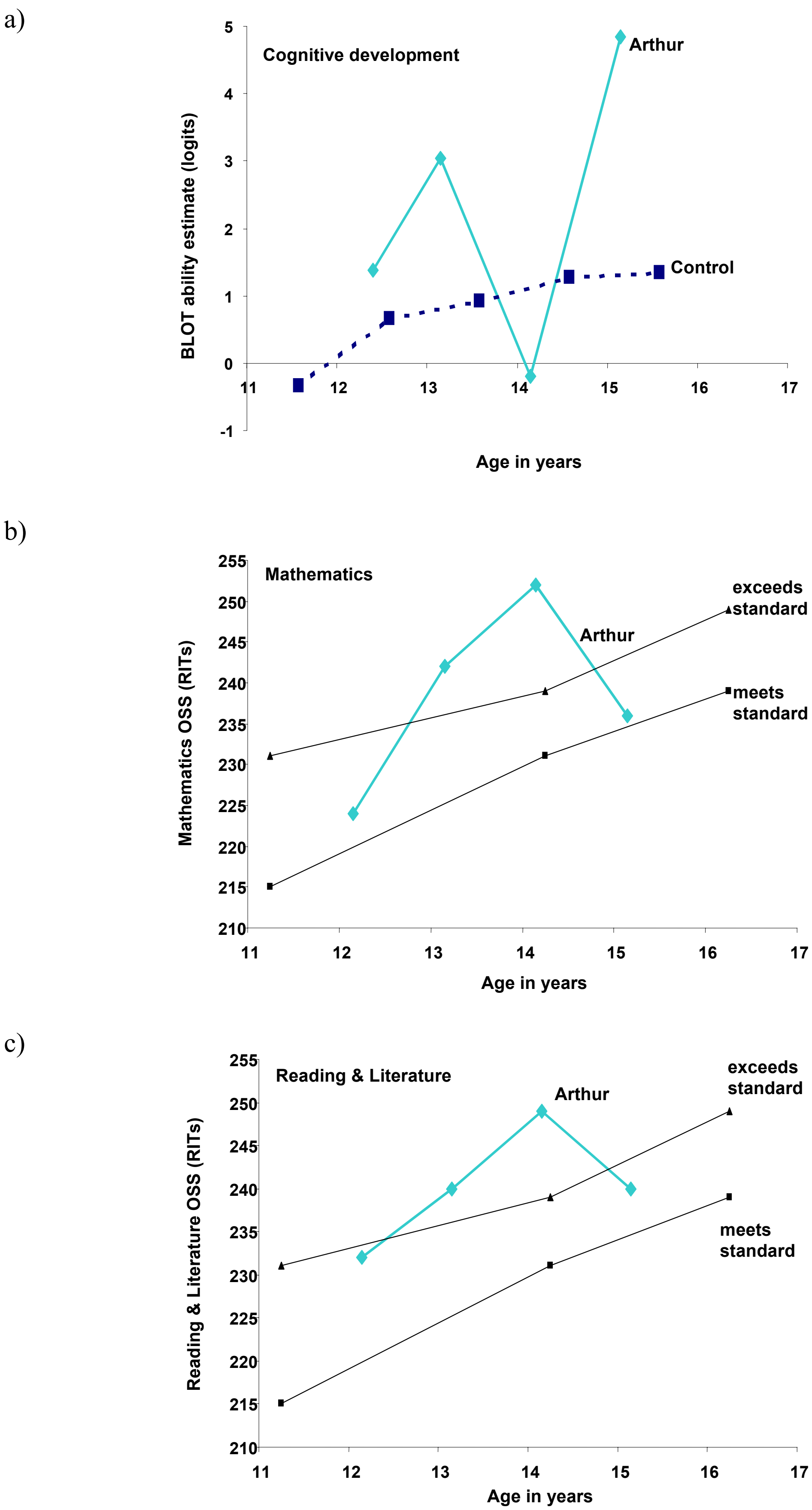


Figure 7.4 Cognitive development and achievement profiles for Arthur

Arthur exceeded the 8th grade standard for Mathematics and Reading & Literature. However, his results for 9th grade were lower, closer to the benchmark for meeting the standard. His Mathematics OSS had increased by 12 RITs (224, 242, 252, 236) and his Reading & Literature OSS by 8 RITs (232, 240, 249, 240) by the end of the study. Arthur's 8th grade Science OSS was 244, indicative of meeting the standard.

Arthur's fit statistics for the BLOT test of 2001 were located within the range of acceptable values, that indicated that his score was probably genuine. It appears that he had not been guessing the answers, nor filling in the bubbles on the Scantron sheet at random. One hypothesis to explain his low score of 2001 could be that Arthur's cognitive development was in fact inhibited in some way during 8th grade. Perhaps it is not a coincidence that his Mathematics and Reading & Literature achievement show a corresponding downturn the following year. Indeed, the mean results for Arthur's male peers also showed the cognitive downturn in 8th grade, followed a year later by a fall in achievement scores. This pattern of results might be evidence of a delayed effect of cognitive development on student achievement. A lag between increased cognitive development and improved achievement was common in CASE interventions implemented in the UK. In this study, the lag is associated with a cognitive development/achievement drop rather than a gain.

Ron

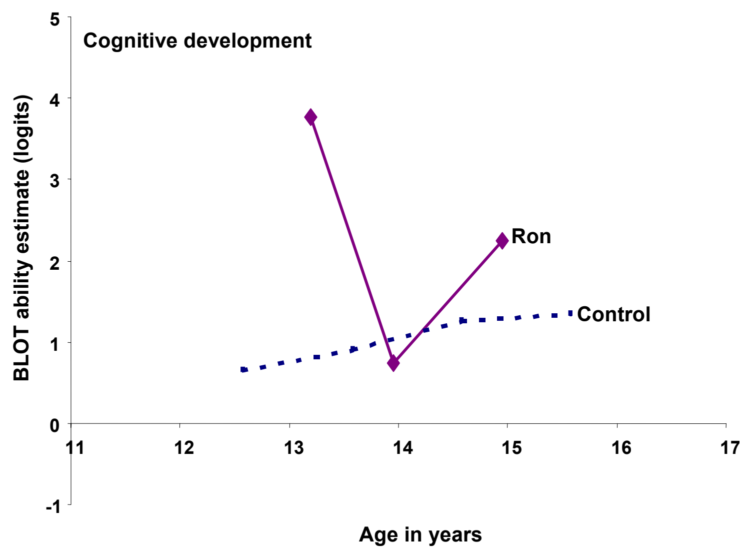
Ron was one of the more able students in Cohort C, aged 13 years 2 months at the start of the STEP intervention. His parents' level of education was 3 (some college education). Twelve CASE activities were delivered to Ron and his peers in Cohort C in 8th and 9th grades. Ron's cognitive developmental level was estimated to be at the mature formal operational stage in September 1999. He was absent for the BLOT test of 2002. Figure 7.5a reveals that Ron's pattern of cognitive development shows the fall during 8th grade that was typical of male students in his cohort.

Table 7.6 *Cognitive development profile of Ron*

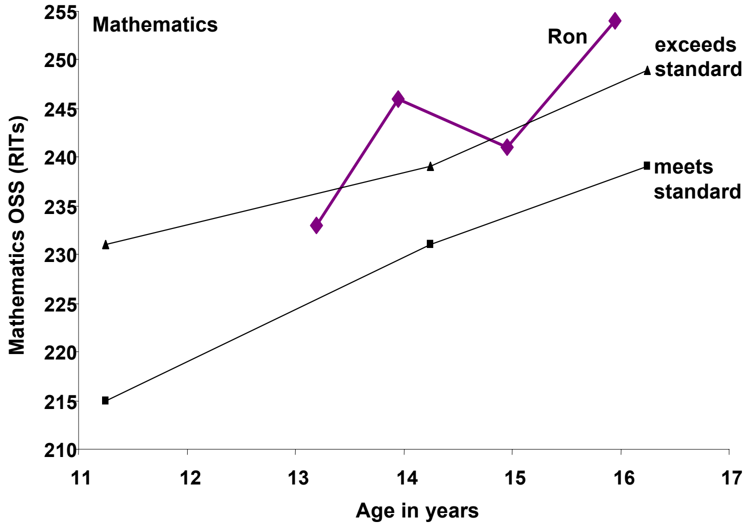
Change in BLOT ability estimate (logits)	Additional correctly answered BLOT questions	New formal operations acquired
i) 3.77 to 0.74	None gained	N/A
ii) 0.74 to 2.25	None gained	N/A
iii) N/A (absent)	N/A	N/A

Ron's Mathematics results (233, 246, 241, 254 RITs), shown in Figure 7.5b, improved from meeting, to exceeding, the state standard by the end of the study. His Reading & Literature scores (244, 246, 249, 248 RITs) were all in excess of the state standard. Ron's performance in Reading & Literature was atypical, as females usually out-perform their male peers in this subject in the Molalla River School District. In the state-mandated science tests, his 8th grade OSS of 245 RITs met the standard, and his 10th grade result (253 RITs) exceeded it. Like Arthur, Ron's results for cognitive development and mathematics could also be interpreted as providing evidence of a delayed effect of cognitive development on student achievement.

a)



b)



c)

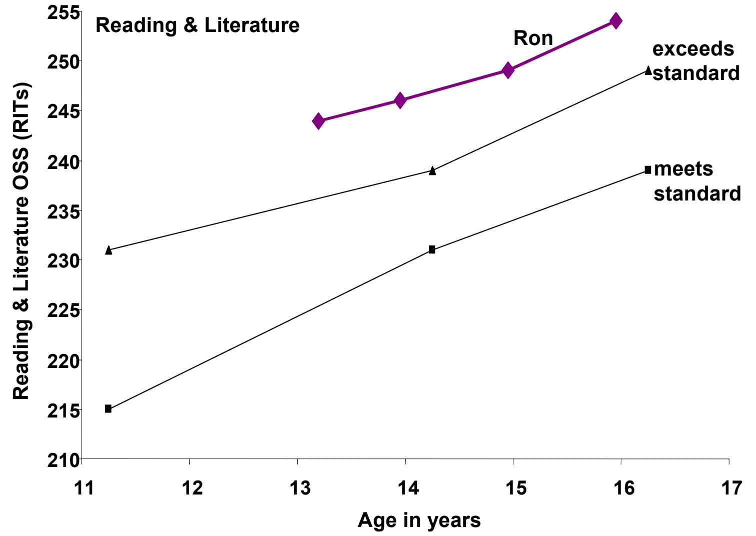


Figure 7.5 Cognitive development and achievement profiles for Ron

Discussion

In general, it is highly likely that children will acquire the mental skills associated with early formal operational thinking before those related to the more advanced mature formal operations, shown higher in the BLOT scale in Table 7.1. However, the relationship between children's cognitive development and the acquisition of particular new formal operational skills was found to be highly variable. Vicky, for example, achieved a correct answer to the most difficult BLOT question, # 32, at age 12+/13+ (her outfit mean square was 1.36 in 2000, slightly beyond the acceptable range of 0.75-1.30). Whereas it was only at age 14+/15+ that she correctly answered Question 7, associated with the acquisition of the relatively more simple operation of correlations (outfit mean square 0.91 in 2002).

The five individual studies provide a taste of the diversity found in the patterns of cognitive development among the children in this study. Figures 5.3, 6.1, and 6.9 clearly illustrate the general trends that occurred in the cognitive development and student achievement of each cohort of students over the 32-months of the study. In contrast, these vignettes show that such trends can be seen in the results of individual students, and are not just artifacts of the statistical treatment of the data. The vignettes also reveal what actually changed for these children in terms of the new mental tools that each had acquired over the period of the STEP intervention.

CHAPTER 8

STEP Teacher Questionnaire

A survey of teacher satisfaction with the STEP intervention was conducted in the Spring of 2003 as part of the evaluation process of the STEP. Dr. Sandra Pellens, the Director of Instruction of the Molalla River School District, conducted the survey and was responsible for its distribution and completion by the STEP teachers. Dr. Trevor Bond and I suggested topics to be included in the questionnaire, which is included as Appendix 2.

Twelve teachers participated in the STEP between 1999 and 2002. These teachers are identified by pseudonyms in Table 8.1 to ensure anonymity. Only ten of the twelve STEP teachers responded to the survey. The two who did not respond were Susan, a newly hired 9th grade teacher who taught STEP for the first time to Cohort B in 2001/2002, and Bill, an 8th grade teacher who took part in the initial training in 1999/2000 but retired at the end of that school year. Colleagues reported that Bill did not teach any of the CASE activities, but had Nancy deliver them to his classes. Alan, who had taught the STEP for one year, completed questions 1 to 11 only of the questionnaire. The average length of time that the STEP teachers had taught science was 10.7 years, and only three had less than five years experience. The 6th grade teachers had elementary education certification, whereas the teachers of the higher grades had Integrated Science and/or Biology secondary education qualifications.

Table 8.1
STEP teachers

	STEP cohort taught				
Pseudonym	**99-00**	**00-01**	**01- 02**	**Years of teaching science**	**Response to survey**
Carly	A			11	✓
Lynne	A			3	✓
Julie	B	A		15	✓
Steve	B	A		2	✓
Nancy	C	B	A	4	✓
George	C	B	A	13	✓
Tom		C	B	16	✓
Paul		C	B	24	✓
Joan		C	B	9	✓
Alan			B	10	✓
Susan			B	Undeclared	
Bill	C			Undeclared, retired in 2000	

Results of the survey

The STEP teachers' responses to the evaluative questions 4 to 7 are shown in Table 8.2. It is interesting to note that the same two teachers who reported that they did not enjoy teaching the STEP lessons also indicated that they did not believe that their students gained much benefit from the intervention, nor did they feel that STEP had affected their usual teaching style. These were the 8th grade science teachers, Nancy and George. In response to Question 24, 'Do you think STEP should be more generally used in schools?', Nancy replied: "No, we do less canned, more inquiry based lessons than that in regular class." She also complained that the "equipment was shoddy or not provided - you need tools to make it - not readily available to order with the program, which it should be." George commented that "the ways the lessons are written are weak" and "No, I am not particularly fond of the activities. I can approach concepts with better activities."

Table 8.2
Results for questions 4 - 7

Question	A little 1	2	3	4	a lot 5
4: Do you enjoy teaching STEP lessons?	Nancy	George	Carly, Lynne, Julie, Steve, Joan, Alan	Tom, Paul	0
5: Do you believe that students gain benefit from STEP ?	Nancy	George	Lynne, Alan	Carly, Julie, Steve, Tom, Paul, Joan	0
6: Do you feel that using STEP has affected your usual teaching style in other lessons?	Nancy, Carly, Joan	George, Julie, Steve	Lynne, Tom	Paul, Alan	0
7: Did you encounter problems with teaching STEP?	Tom, Alan	0	Lynne, George, Joan	Carly, Julie, Steve, Paul	Nancy

Nancy and George stated that using the STEP had not affected their usual teaching styles. Both reported that they did not use STEP language in other science lessons, nor any components of the typical STEP lesson plan, and they did not notice students making connections between STEP and other science lessons. They also stated that given the choice they would not wish to continue with STEP, and did not feel the program should be more generally used in schools. Nancy and George taught Cohort B in 2000-2001, the year in which the mean BLOT ability estimate of that cohort showed a significant downturn of large effect size (see Figure 5.3 C in Chapter 5). It is tempting to ask if there might be a relationship between these teachers' attitude to the STEP (and their teaching strategies in general) and the progress of their students.

Nancy took maternity leave in the spring of 2001 and was replaced by a substitute teacher, whose only resource-person for the STEP was George. No significant difference was found between the cognitive development of Cohort B students in Nancy and George's classes that year, in spite of Nancy's classes receiving eight CASE activities to George's three. Could it be that the way in which these teachers delivered the CASE lessons was sub-optimal? In the absence of Nancy, George was the person responsible for the administration of the BLOT test to the 8th grade classes in May 2001. If George's negative attitude to the STEP also extended to the BLOT, then the cognitive test might have been presented to the students in a manner that did not encourage them to take it seriously. Indeed, the BLOT answer sheets from George's classes that year were misplaced, and I did not receive the results until the spring of 2002.

In contrast, the teachers of the other grades (6th, 7th and 9th) had more positive things to say about the STEP intervention in their responses to the survey. For example, answers to Question 8: 'What aspects of STEP teaching did you most enjoy?' revealed general teacher satisfaction with the CASE activities:

Lynne:	"actual experiments - what the kids enjoyed"
Julie:	"The concepts behind STEP."
Steve:	"hands-on. I feel the kids learned the most from STEP when using hands-on activities."
Tom:	"I like that the lessons put responsibility on the students to clarify their own thinking/communicating."
Paul:	"Inquiry - makes kids think."
Joan:	"hands-on activities; keeping the students actively engaged is an excellent learning opportunity"

The responses of Julie, Tom and Paul reveal understanding of the theoretical concepts on which the CASE program is based. Whereas, those of Lynne, Steve and Joan suggest that these teachers might have missed the point of the intervention.

Lack of time and equipment were commonly cited in responses to Question 9 as the least enjoyable aspects of teaching STEP, so much so that two 9th grade teachers admitted to skipping some STEP lessons due to pressure of time and missing equipment. Responses to Question 9 included:

Carly:	"Time away from the curriculum."
Lynne:	"prep time"
Julie:	"The STEP workbook was not interesting to the students - looking at pictures and answering questions."
Steve:	"Trying to relate confusing language & concepts to the students."
Tom:	"None - unless it was not having equipment for the electricity activities."
Paul:	"Some activities were too simple or too low."
Joan:	"finding time to do them, prep work, & they are 'misplaced'/ no flow to the curriculum (even though it was supposed to be that way, I didn't enjoy doing that)"

Tom and Paul were more motivated and supportive of the STEP than other Molalla High School teachers. Their responses indicate that they did not regard the STEP as an intrusive and time consuming addition to the curriculum, as was the view of many of their colleagues.

Several answers to Question 10 concerning the areas of science in which students seem to gain most benefit from STEP alluded to logical thinking and/or the concept of controlling variables:

Carly: "Scientific method"

Lynne: "Logical thinking"

Julie: "All areas"

Steve: "Classification. Variables/Fair Test"

Tom: "Students are helped to develop a good sense of experimental process ('fair test'), logical arguments, communicating clearly/concisely."

Paul: "All equally"

Joan: "lab activities & anything that requires higher level thought processes"

One of the professional development goals of the STEP was enhancement of the teachers' understanding of the relationships between student abilities and the processes of scientific inquiry. The general agreement that the project was beneficial in helping students to develop scientific thinking skills suggests that this aspect of the teachers' professional growth was successful.

There was an interesting variety of answers to Question 11 that asked whether some students appear to benefit more than others from the STEP. It appears that these teachers judged the benefits of the STEP based on students' performance in those lessons, rather than taking a longer-term view of their pupils' achievement:

Carly: "Yes - those who are capable of more higher-level thinking"

Lynne: "Yes, those with the ability to think logically already"

Julie: "Yes, some students who are beyond concrete operational thinking understood activities. Many students (IEPs, ESL, lower level students) didn't get it."

Steve: "Yes. The students who try and struggle benefit the most. Some students give up trying."

Tom: "Not that I've noticed - some students 'get it' more quickly - but that is always the case."

Paul: "Yes, the ones that actually solve the problem. Others just follow along."

Joan: "Those that learn by doing rather than structured book work benefit more."

Joan's answers to Questions 10 and 11 indicate that she values the hands-on aspect of STEP activities, which might suggest a misunderstanding of the CASE rationale. Carly, Lynne, and Julie share the feeling that it might be the more able who gain the most benefit, whereas Paul and Steve see engagement with the problem to be an important element. It is only Tom who sees that CASE might benefit a broad range of students.

Although the majority of STEP teachers felt that the intervention had not greatly affected their usual teaching style, three gave examples of how it had done so in their answers to Question 12:

Carly: "Surprising events"

Tom: "The questioning process - getting students to communicate their thinking."

Paul: "I have included more inquiry activities. Less direction."

Tom's answer, in particular, indicates an understanding of the metacognitive aspects of the CASE activities.

Some teachers had noticed students making connections between STEP lessons and other science lessons (Question 13):

Carly: "Yes - fair test"

Julie: "When we apply the scientific method in other lessons and when they did their science project they remembered controlling variables."

Steve: "Yes. Only when referring to variables & fair test."

Tom: “Yes. We’ve integrated the STEP lessons to reinforce the curriculum. So, for example, ‘the pressure lab’ - students have referred to the results when trying to explain/explore other pressure labs/demos.”

Examples of their using STEP language in other science lessons, given by teachers in answer to Question 14, included:

Julie: “Yes, we talk about variables, input and output variables, controls *etc*.”

Steve: “Yes. See above # 18." (“Yes. Only when referring to variables & fair test”)

Tom: “The STEP language I use most is: ‘fair test’, variables, values, relationships, compound variables...”

Paul: “Fair test, variables, control, *etc.* - we use all year.”

Joan: “Yes, variables, input & output, *etc*.”

The answers to Questions 13 and 14 indicate that, in general, these teachers regarded CASE activities 1-5 (Variables) as being the key components of the STEP. Indeed, the teachers repeated these five activities at the start of each academic year in order to reinforce the concept of a fair test.

Teachers were also asked whether any components of the typical STEP lesson plan were used in other science lessons (*e.g.* concrete preparation/surprising event/group discussion/bridging, *etc.)*:

Carly: “Always discuss fair tests. Discuss why a surprising event occurred.”

Julie: “No. Did these things before STEP.”

Steve: “I have not made the correlation between STEP lesson components & my own lesson plans.”

Tom:	"Yes - these are not unique to STEP - they represent good teaching practice."
Paul:	"All of them at one time or another"
Joan:	"Yes, but I don't label each individual event. It comes naturally after yrs. teaching ☺"

Steve, a second year teacher, was the only person who stated that he had not incorporated any of the CASE lesson components into his own teaching plans. The others, with many more years teaching experience, regarded the CASE lesson components as characteristic of good teaching practice.

Examples of problems experienced by the teachers with the STEP included time constraints, lack of equipment and cultural issues (Question 15):

Lynne:	"Lack of materials, running out of time"
Julie:	"Difference in language, lack of materials"
Steve:	"Confusing language"
Tom:	"Some 'cross cultural problems' (the worksheet with the lemon concentration on the back; the pressure worksheet with the jeep crushing someone on the back ..."
Paul:	"Too low a level. Not enough equipment or time."
Joan:	"None really"

There are no laboratory assistants/managers in the Molalla schools, and STEP teachers had to construct their own apparatus for the CASE activities. It appears that insufficient time allowance was made for the teachers to prepare the new materials. The references to the language of the CASE activities indicate the need for teachers to adapt the worksheets for an American audience. Paul's comment about some CASE activities being at too low a level for his 9th graders suggests that the high school teachers might need to review the role of the STEP in their curriculum.

Question 16 asked the teachers about the training they received to prepare them for teaching STEP. Their answers highlight the fact that they received little support after the initial training sessions in 1999/2000:

Carly:	"2 day training workshop"
Lynne:	"district support"
Julie:	"Several workshops"
Steve:	"advice from other teachers"
George:	"In service days, monthly meetings"
Nancy:	"workshops provided by the district"
Tom:	"Several hours of training followed by regular meetings during the first year."
Paul:	"A few sessions"
Joan:	"Workshops"

Question 17 'Would you personally continue to use STEP?' elicited a variety of responses:

Carly:	"No - I believe that the ideas of STEP are very important & necessary for those without extensive training in the teaching of science. Their ideas are things well-trained science teachers already incorporate into their lessons."
Julie:	"Some of the lessons, 1-5, classification"
Steve:	"only selected lessons - with limited class time for curriculum, I can only justify a few lessons"
Tom:	"Yes. It compliments the program nicely and effectively."
Paul:	"Yes, but I have adapted most of them."
Joan:	"Probably not unless asked to do so."

It is interesting to speculate whether some of these teachers would have been more positive about continuing with the STEP if they had experienced more professional development and in-class support. Carly's comment that well trained science teachers already incorporate similar strategies into their lessons is indicative of a superficial understanding of the CASE rationale.

Question 18 asked the teachers' opinion on whether STEP should be more generally used in schools:

Carly:	"In elementary schools if it could be geared toward that age level or in schools where teachers are not extensively trained."
Lynne:	"to help with logical thinking"
Julie:	"good introduction"
Steve:	"Yes, with room for modification & use of only selected lessons."
Tom:	"The STEP activities have obvious value and could be an effective part of any science program."
Paul:	"The inquiry idea is good, but STEP is only a small part"

The teachers varied in their opinions about the best grade levels to target with the STEP, although most regarded the program as more suited to the middle school years:

Carly:	"5 - 6, before more formal science classes begin"
Lynne:	"8-12, ability to think outside the box"
Julie:	"7, 8th grade"
Steve:	"7-12. I would begin in the 7th grade when students are developing their processing skills."
Tom:	"6 - 8"
Paul:	"I do think the STEP activities should be moved 'down' to the middle school."
Joan:	"Younger, I suppose. To keep track of progress as they become older students."

The last two questions (20 & 21) in the survey referred respectively to the teachers' and students' experiences with the BLOT cognitive development test:

Carly: "Confusion on the students part." "Frustration unless they were very analytical."

Lynne: "Understanding issues w/students." "Another test?" "This one again!"

Julie: "None." "Not much reaction except poor readers had a difficult time."

Steve: "No problems, just another standardized test to them."

George: "None." "Confused. Questions worded strange. After the 2nd time they get used to them, and usually just start bubbling."

Nancy: "w/groans - 8th grades have already taken it 2x"

Tom: "None I recall - except, again, finding the time!" "Students were often overwhelmed - the tests are language "dense" and challenging to students."

Paul: "Understanding the question." "Think it is easy but don't do well. Repeat times - same idea."

Joan: "Okay"

Discussion

The teachers' responses to the survey provided an important evaluation of the STEP program, and a basis for making recommendations regarding the future implementation of the intervention. While the responses to the questionnaire might be seen to reflect the professionalism of these teachers, it is clear that the intervention was delivered in a field setting that was less than optimal. The training and support available in Molalla was very different from that typically provided for CASE programs in the UK, where teachers had access to the CASE professional development network and materials.

Eight of the ten teachers who responded to the survey expressed general satisfaction with the CASE activities. However, at least six teachers struggled with the issues of lack of time for the preparation and teaching of the STEP activities, and/or experienced frustration over inadequate or missing equipment. There are no laboratory managers/assistants in the Molalla schools, and the construction of the equipment for the CASE lessons was therefore the responsibility of the teachers, working from instructions in the *Thinking Science* (Adey, Shayer & Yates, 1991) teachers' manual. One recommendation to the administrators of the Molalla River School District would be to provide technical assistance for the STEP teachers. A well-maintained materials bank would alleviate the teachers' problems associated with inadequate equipment.

Time constraints were cited by one 7th grade teacher as the reason why only a few CASE lessons could be justified. Furthermore, a 9th grade teacher reported that she felt that there was no flow to the curriculum, and suggested that some CASE activities were misplaced. Perhaps there is a need for the teachers to review the science curricula of the Middle and High schools, in order to ascertain whether a different pattern of the CASE activities might better fit the program for each grade. If such a review took into account the Oregon State Science Standards, it might also reveal any overlap in the content of the science concepts covered each year. The pressure on teachers to cover content is enormous in a standards-based system such as that found in Oregon. In the light of the benefits of STEP to students, it would be a pity if such time constraints caused the Molalla River teachers to abandon the intervention.

Four teachers expressed concern about the language used in the *Thinking Science* curriculum materials, that were written for British children. Perhaps there is need for teachers to adapt the student worksheets more specifically for an American audience? One 9th grade teacher (Paul) regarded some of the activities as being at too low a level for his high school students. This suggests that a review of the specific CASE activities taught to each grade level might be in order to ensure a better match between cognitive level and the curriculum demand of the CASE activities.

The results of the STEP teacher questionnaire revealed that some teachers had misconceptions concerning the rationale behind the intervention. To the question of what parts of STEP did they most enjoy teaching, Steve and Joan reported that they liked the "hands-on" aspects of the lessons, and Joan further commented that the lessons kept her students engaged. These answers suggest a low level of understanding of the cognitive nature of the intervention. In response to the question about what aspects of STEP most benefit students, Steve selected classification, variables, and the fair test, and Joan the laboratory activities. It appears that both of these teachers have not seen beyond the application of the STEP lessons to the content and skills of their regular science curriculum. Indeed, their comments suggest that these teachers had missed the point of the whole intervention.

However, some of the teachers' comments revealed that their practices had changed through using the CASE activities. Examples of how their teaching styles had changed were provided by three teachers, while five described instances of how they now use STEP language in other science lessons. These claims are heartening, and suggest that some positive changes associated with the STEP are evident among the teachers, as well as the students, of the Molalla River School District.

CHAPTER 9

Conclusions

This chapter presents a summary of the results, conclusions, and implications of the study. Analysis of the results of the Molalla River sample has revealed information relating to the mode of delivery of the CASE strategies that has not been described previously. The discussion that follows confirms that some of the benefits of the CASE intervention, described in research conducted in other countries, also apply to the STEP. However, these benefits do not appear to be as impressive as those claimed for the original CASE research in the UK.

Given that the STEP teachers received little professional development, and did not deliver the entire intervention program, it is not surprising that STEP did not yield achievement results as strong as in the original CASE projects in the UK. Furthermore, the CASE literature reports a gap of several years between improvement in cognitive development and significantly better achievement scores (Adey & Shayer, 1990, 1991, 1994). Perhaps it was too early, after only 32 months, to find evidence of far-transfer effects in the achievement results of the STEP students, such as were found in Britain.

If this research had been a funded project to investigate the impact of the CASE intervention in the Molalla School District, the STEP would have been based on a more thoroughly implemented and systematic version of what actually took place. A fully supported teacher development program would have been undertaken, as well as non-project control classes set up in the Molalla schools. Because the Molalla School District did not allow the setting up of parallel control and experimental classes, it was necessary to use comparison data from a pre-intervention cross-sectional profile, and data from carefully selected school districts nearby. Conclusions based on cross-sectional controls and comparison with the non-project schools, as were employed in the Molalla River study, must necessarily be regarded as more tentative than conclusions based on within-school controls, such as those used in the original CASE research.

Summary of Results

Statistically significant overall effects were detected for cognitive change in Cohorts B and C compared with the cross-sectional control. The gain over the control for Cohort B was evident in three separate analyses: pooled results of males and females tested with BLOT on a minimum of two occasions, females analysed separately, and students tested on all four occasions. The statistically significant overall gain for Cohort C over the control was based on the BLOT results of students tested on all four occasions. No overall statistically significant difference was evident between the cognitive change of Cohort A and that of the cross-sectional control.

The overall results for cognitive change suggest that it was the students who began the intervention in 6th and 7th grades who apparently gained the most benefit from the intervention. It is not possible to determine from the results of this study why the beneficial cognitive effects of STEP did not apply to the children who started the STEP in 6th grade. Possible explanations might include the mismatch of the STEP with the children's developmental level, lack of effective long-term professional development for the STEP teachers, and the number (17/30) and nature of CASE activities covered.

Statistically significant overall gains in Mathematics achievement were made by Cohorts A and B over the cross-sectional control. This effect was also seen when results were analyzed separately for Cohort A males and females. An overall gain over the control was also detected in the separately analyzed Reading & Literature results of Cohort B females.

In terms of achievement in the state-mandated Mathematics test, the overall results suggest that it was the students in Cohorts A and B who began the STEP in 6th and 7th grade who gained most benefit from the intervention. It is interesting that no achievement gains were found for Cohort C, because its sub-set of students that was tested on all four occasions, showed enhancement of cognitive development, but not of Mathematics nor Reading & Literature achievement. Cohort C received only 13/30 CASE activities, delivered over 8th and 9th grades. Perhaps this was enough to enhance cognitive change for some individuals, but insufficient to enable the transfer of new cognitive skills to new contexts.

No statistically significant overall effects were detected when the student achievement of STEP participants was compared with two non-project school districts. Data from these carefully selected comparison schools were collected in good faith, and it would be less than open to fail to disclose them at this point. However, in hindsight, there were many unexplored differences between these schools and those of the Molalla River School District (for example, school ethos, professional development emphasis, and curriculum offerings) that would be likely to influence student achievement. Conclusions based on the between-schools comparisons must, therefore, remain tentative.

Research goals

Goal 1: to investigate whether cognitive development was enhanced by the STEP intervention in the Molalla River School District.

Specific research questions:

1.1. Has the cognitive level of STEP students changed over the period of the study?

Significant improvement took place in all three cohorts of STEP students over the period of the study. Although some students who started with a low BLOT ability estimate in 1999 achieved a similar score in 2002, many more of their peers made gains of up to 5.0 logits over the period of the study. This represents a gain of more than six standard deviations for these individuals. It seems logical that cognitive change should indeed take place over a 32-month period during adolescence. A more important question, therefore, is whether more cognitive change took place in the STEP students than in the control.

1.2. How is the pattern of cognitive change of STEP participants different from that of the Molalla River school sample prior to the intervention?

Figures 5.4 a, b, c in Chapter 5, which display mean BLOT ability estimate plotted against average age in years, revealed different patterns of cognitive change for the control sample and the three cohorts of STEP students. The control profile showed a period of rapid change from average age 11+ to 12+ years, followed by a plateau in the mid-teenage years, a pattern of cognitive change in accord with data for more than 10,000 British children (Shayer, Küchermann and Wylam, 1976).

The three cohorts of STEP students were respectively aged 11+, 12+ and 13+ years at the start of the intervention. Cohort A showed the most similarity with the control profile, having a cognitive growth spurt at average age 11+ to 12+, followed by a period of slower growth. Data collection started for both Cohorts B and C after their early adolescent growth spurts, at average age 12+ and 13+ respectively. The proximity of the mean BLOT starting values of Cohorts B and C to the control profile suggests that the cognitive developmental levels of these students prior to the intervention were similar to those of the control.

Statistically significant overall effects compared with the cross-sectional control were detected for cognitive change in Cohorts B and C. This gain over the control for Cohort B was evident in three separate analyses: pooled results of males and females tested with BLOT on a minimum of two occasions, females analysed separately, and students tested on all four occasions. The statistically significant overall gain for Cohort C over the control was based on the BLOT results of students tested on all four occasions. No overall difference was evident between the cognitive change of Cohort A and that of the cross-sectional control.

However, it was in the last year of the study that the most striking differences were seen between the cognitive change of the STEP and control students. Both Cohorts B and C showed statistically significant gains of large effect size over the control sample in the last year of the study. Further evidence to support the idea that STEP participants show different patterns of cognitive development from the control profile came from the estimation of the Piagetian levels of the control profile and the STEP students in May 2002. The percentage of students showing mature formal operational thinking was greater compared with the control in all three STEP cohorts by the end of the study.

These results suggest that students in Cohorts B and C, on average, were indeed more cognitively developed, according to the BLOT, than their age peers in the control sample by the end of the STEP intervention. Enhancement of cognitive development was also found in the CASE interventions implemented in the UK, Pakistan, Denmark, Finland, Malawi, and Australia.

1.3 What is the relationship between cognitive change and level of parent education for STEP students?

There was no significant relationship between the cognitive change of STEP students and the level of education of their parents. Some students from the whole range of level of parent education ('not high school graduate' to 'advanced degree') had high gains in BLOT estimates. In US terms, the range of level of parent education in Molalla River was relatively narrow, with few individuals at levels 1 and 5 ('not HS graduate' and 'advanced degree'), reflecting the demographics of this agricultural community (US census 2000). Given that the population of Molalla has a lower percentage of parents with advanced degrees than either the state as a whole or the USA (US census 2000), these results might indicate that the STEP intervention has potential as a program to assist school populations in middle to low socioeconomic settings. If a relationship between cognitive change and level of parent education had existed in the sample prior to the STEP, then one could propose that the intervention might have redressed the effects of any such imbalance among these children.

1.4 What is the relationship between cognitive change and cognitive developmental level at the start of the STEP intervention?

An important question for this study was whether cognitive change would take place across the whole spectrum of starting ability. Perhaps it would be only students with more ground to make up who would show improved scores; or, alternatively, the more cognitively developed students who were more "ready" might be the ones to show most change. It was therefore interesting that no significant relationship was found between starting ability and change in cognitive level. High gains were made by some STEP participants across the full range of starting ability, irrespective of their cognitive levels prior to the intervention. These results confirm those of the UK and Australian CASE studies, which found that starting level is not a significant factor in cognitive growth.

Further support for the notion that the full developmental range can benefit from the STEP intervention came from analysis of the relative means of annual change in BLOT ability estimates. Little bias towards the starting level was found for both Cohorts A and C in the first year of the intervention, followed by a trend of less able students improving faster during the second year. By the third year, it was the students who started at a higher level who made more improvement. The pattern was quite different for Cohort B: higher ability students improved faster in the first year, followed by their lower ability peers making greater gains in the second and third years of the intervention. Starting the STEP in 7th grade might target students at a critical age for which the intervention had a delayed effect on lower achievers. The analysis of the relative means of the BLOT ability estimates of each cohort provides a second line of evidence that gains in cognitive developmental level were made across the entire range of starting ability of the STEP participants.

1.5 What is the relationship between cognitive change and the age at which the STEP intervention commenced?

No statistically significant relationship was revealed between the age at which the

STEP intervention commenced and the change in the BLOT ability estimate from 1999 to 2002. Some students across the entire age range of 11+ to 14+ years made substantial cognitive gains over the period of the intervention. These results add further weight to the argument that the STEP intervention appears to benefit a broad spectrum of students, refuting the idea that there might be a simple relationship between age and cognitive development.

1.6 What is the pattern of cognitive change for individual students?

Analysis of the BLOT results for individual students showed that there was remarkable variation in the pattern of cognitive change for individual STEP students. Indeed, when the BLOT ability estimates were plotted for those students who were tested on all four occasions, the lines for each individual appeared to be unique. Unique development patterns were also found for the children in our Australian CASE study (Endler & Bond, 2001).

The amount of change in the BLOT ability estimate per year also varied considerably from person to person in the three STEP cohorts. There was a general trend of increasing numbers of individuals achieving higher ability estimates over the period of the program, suggesting that students were indeed developing cognitively. This trend might well be explained by early adolescent development in Cohort A, as students mature from 11+ to 13+ years. However, the trend is also found in Cohorts B and C, suggesting that factors other than mere adolescence might be implicated in the cognitive growth of these older students.

1.7 How is the pattern of cognitive change different for males and females?

Different profiles of cognitive change were found for male and female students. In the control group, the trend was for control females to show more or less steady improvement from age 11+ to 14+ years, followed by a decline. Whereas the trend for males was a growth spurt from 11+ to 12+ years, a plateau from 12+ to 13+, followed by a period of slower growth.

No significant difference was found between the rates of change in BLOT ability estimate of females in Cohorts A and C and the control data for females. In contrast, Cohort B females showed a statistically significant overall gain over the control. There were also significant gains for Cohort B females in the first and third years of the STEP intervention, interspersed by a significant loss. The only significant gain over the control sample for males was made by Cohort B in the last year of the intervention, which again followed a significant loss in 2000-2001.

Analysis of the results for female and male STEP students revealed that female students in Cohort B might have gained more benefit from the intervention than their male peers. Furthermore, it suggested that the STEP intervention could be used specifically to target and enhance the cognitive development of 12-year-old girls in the Molalla River School District. Recent TIMSS reports suggested that gender differences were minimal in mathematics achievement in the USA, but pervasive in science, particularly in physics, chemistry and earth science (NCES on-line, 2003). It therefore seems likely that a science-based program, such as STEP, that is claimed to enhance the achievement of girls, might be useful in the current educational climate of the USA.

It is interesting that the results for the male and female STEP students in Cohort B contrast with those of the UK CASE studies, where it was the boys who started the program at age 12+ who were the only sub-group to show significant gains in cognitive level over the control (Adey & Shayer, 1990). The small group of girls in the Australian CASE study also performed better than would be predicted from the UK CASE results. It would appear that the effects of gender on cognitive development are complex, and might be influenced by cultural factors.

1.8 How is the pattern of cognitive change different for students who receive only part of the intervention?

Two subsets of students were identified, each of which had received only part of the STEP intervention. One subset, peers of Cohort A, missed the first year of the intervention. The pattern of cognitive change of these students, who missed STEP in 6th grade, was found to be similar to that of the control profile. These students did not actually miss any content of the STEP, as the CASE activities (1 to 5) taught in 6th grade were repeated the following year.

In contrast, the pattern of cognitive change for the peers of Cohort B who missed the STEP in 8th grade was found to be significantly different from their peers who experienced the whole intervention available to their grade. Arguably, this could be regarded as one of the more important results of the study. A statistically significant gain was found between the cognitive change of the STEP students and their peers in the sub-group from 2001 to 2002, suggesting that missing a year of the STEP had inhibited the cognitive growth of the sub-group. The students in Cohort B experienced either 21 or 16 CASE lessons, depending on their 8th grade teacher, whereas the peers of Cohort B were taught only 12 CASE lessons. This important result raises questions concerning which CASE activities are the most important in promoting cognitive enhancement, and also the role of teacher attitude towards the intervention.

Goal 2: To discover whether cognitive change is accompanied by change in the Mathematics and Reading & Literature achievement of STEP participants.

Specific research questions:

2.1 Has the achievement of STEP participants in Mathematics and Reading & Literature changed over the period of the study?

Regression analyses performed on the results of the 1999 and 2002 state-mandated Mathematics and Reading & Literature tests indicated that significant improvement had taken place in all three cohorts of students over the period of the study. However, in order to determine whether the school achievement of the STEP students was enhanced over the period of the intervention, it is necessary to compare the pre- and post- intervention results.

2.2 How are the STEP participants' patterns of achievement in Mathematics and Reading & Literature different from those of the Molalla River school sample prior to the intervention?

Cohorts A and B made statistically significant overall gains in Mathematics achievement over the Molalla River school sample obtained prior to the intervention. This effect was also seen for males and females in Cohort A, when their results were analyzed separately. Different patterns of Mathematics achievement were detected for the control sample and the three cohorts of STEP students. Cohort A made a significant gain in Mathematics achievement over the control between May 2000 and May 2001. Whereas Cohort B showed significant gains over the control profile in 1999-2000 and 2000-2001, followed by a significant loss the following year. The only significant difference in Mathematics achievement for Cohort C over the control was a loss from 2000-2001. Given these results, it appears that the pattern of Mathematics achievement of students in the three STEP cohorts was indeed different from that of the control sample.

In contrast, the only significant overall gain in Reading & Literature achievement during the period of the STEP intervention occurred for the females of Cohort B over the control, although the pooled results of Cohort B males and females did not show this overall significance. If these results for Mathematics and Reading & Literature can be interpreted as providing some modest evidence that the STEP might have enhanced school achievement, then it would appear that it was the female students of Cohort B who gained most benefit from the intervention.

2.3 What is the relationship between change in achievement level in Mathematics and Reading & Literature and level of parent education for STEP students?

No significant relationship was found between the level of parent education and change in Mathematics OSS or Reading & Literature OSS from 1999-2002. Neither the Mathematics nor the Reading & Literature achievement of these students appears to be associated with the educational level of their parents.

2.4 What Is the relationship between change in achievement level in Mathematics and Reading & Literature and STEP participants' starting achievement levels?

A significant relationship was found between the Mathematics starting level (OSS 1999) and the change in Mathematics OSS from 1999 to 2002, suggesting that a lower starting level is associated with a greater gain in Mathematics achievement. A similar relationship was evident between the starting level for Reading & Literature and the change in Reading & Literature OSS from 1999 to 2002. These results suggest that it is the lower ability students who improved the most, perhaps because they had more ground to make up than their more able peers.

2.5 What is the relationship between change in achievement level in Mathematics and Reading & Literature and age at which the STEP intervention commenced?

No significant relationship was found between the age at which the STEP intervention commenced and the change from 1999-2002 in either the Mathematics OSS or the Reading & Literature OSS.

2.6 How is the pattern of student achievement in Mathematics and Reading & Literature different for males and females?

Comparison of the male and female students in the control sample revealed a similar pattern of Mathematics achievement for the sexes, although females started slightly higher than males, and finished lower. When the Mathematics achievement of STEP students was compared with the control profile, a statistically significant overall gain over the control was found for Cohort A males and females. Both males and females of Cohort B showed significant gains over the control in the first and second years of the study, followed by a large loss from 2001 to 2002. There were no significant gains or losses for the females or males of Cohort C over the control.

A further comparison was made between the Mathematics achievement of male and female students within each particular cohort. The pattern of Mathematics achievement for the males and females of all STEP cohorts was seen to be very similar, and no significant difference was found in any of their annual slopes of change. However, Cohort B males consistently scored 5 to 10 RITs higher than their female peers.

Comparison of the Reading & Literature achievement of male and female control students also revealed similar patterns of cognitive development for the sexes. There was little difference between males and females at mean age 10+ years, however, after this point females slightly outperformed their male peers. The growth spurt for both females and males was at mean age 12+ to 13+. When the Reading & Literature achievement of female STEP students was compared with the control profile, Cohort B females showed an overall statistically significant gain over the control, with a significant gain from 1999-2000, followed by a significant loss from 2001 to 2002. Cohort C females showed a statistically significant overall loss against the control. No overall gains in Reading & Literature achievement were evident for male STEP participants. However, males in Cohort A made a significant gain over the control sample in the second year, and Cohort B males showed a significant gain from 1999-2000, followed by a loss in 2001 to 2002.

The pattern of Reading & Literature achievement was also compared for males and females within cohorts. No significant difference was found between the annual slope of change in Reading & Literature achievement of males and females in any cohort. The patterns for the sexes in Cohort A were seen to be very similar, although the mean score for females was 3 to 4 RITs above those of their male peers. A similar trend was evident for both Cohorts A and C, where females consistently outperformed males. Cohort B showed a different trend, with males generally outperforming their female peers.

If the STEP intervention does indeed enhance school achievement, then its effects on the sexes might be different, given the patterns of Mathematics and Reading & Literature achievement detected for male and female STEP students in this study. The results for the control profile indicated that boys in the Molalla School District generally outperformed girls in Mathematics, whereas girls often outperformed boys in Reading & Literature. The results of the Progress in International Reading Literacy Study (PIRLS) also showed a general trend for the achievement of girls to be higher than boys in the countries that participated in the study (Mullis, Martin, Gonzales & Kennedy, 2003). The significant overall gain made by STEP females in Cohort A in Mathematics might have redressed to some extent the gender imbalance in achievement in this subject in the Molalla School District.

2.7 How is the pattern of student achievement in Mathematics and Reading & Literature different for children who receive only part of the intervention?

The peers of Cohort A, who had missed STEP in 6th grade, started out at a higher mean level in Mathematics than STEP students who had completed one year of the intervention. However, it appeared that STEP students later closed this achievement gap and overtook the latecomers by the end of 2002. Indeed, the STEP students in Cohort A experienced a significant gain over this sub-group in the third year of the study, suggesting that missing a year of the intervention might have inhibited the Mathematics achievement of their peers. Furthermore, the latecomers to STEP showed a significant gain over the control sample in the last year of the intervention. These results might indicate that the STEP intervention had enhanced the Mathematics achievement of these latecomers, though not to the same extent as their peers who experienced the whole of the intervention available to their grade.

The second subset of students that experienced only part of the STEP intervention comprised peers of Cohort B whose 8th grade teachers had opted out of the STEP in the 2000-2001 school year. When the mean change in Mathematics OSS of these students who missed the STEP in 8th grade was compared with that of their peers in Cohort B, no significant difference was found in the first year, when both groups were receiving the STEP intervention. However, in the second year when the groups received different treatment, Cohort B showed a significant gain over the students who missed the intervention in that year. This might indicate that the STEP intervention had enhanced the Mathematics achievement of the STEP students in 2000-2001.

A comparison was also made of the Reading & Literature achievement of peers of Cohorts A and B who missed part of the STEP intervention. However, no significant difference was found between the Reading & Literature achievement of these subsets of students. The results provide no evidence that the STEP intervention had any effect on the Reading & Literature achievement of these students.

2.8 How is the pattern of student achievement of STEP participants in Mathematics and Reading & Literature different from that of peers in comparison schools?

The results of the Molalla River students who experienced the STEP intervention were compared with those of students in two other non-project school districts, selected because of their similar socioeconomic status to Molalla. Overall, no statistically significant effects were found. However, in the state-mandated Mathematics test, a significant gain was made over the Sequoia students by Cohort A from 2000 to 2001, preceded by a loss. Cohort B made significant gains over both Redwood and Sequoia students from May 2000 to May 2001, followed by significant losses. Cohort C showed a loss against Sequoia from 2000 to 2001. In Reading & Literature, Cohort B showed significant losses against Redwood and Sequoia from 2001 to 2002. Conclusions drawn from these data must take into account the many unexplored variables that existed between the Molalla River, Redwood and Sequoia schools.

2.9 What is the relationship between cognitive level and achievement in Mathematics, Reading & Literature and Science for STEP participants?

The consistently strong correlation between the BLOT cognitive ability estimate and the Oregon State Scores in Science, Mathematics, and Reading & Literature could be interpreted as evidence that cognitive level is associated with school achievement for these students. If, as hypothesized, the STEP intervention has indeed enhanced the cognitive development of some of the students who experienced the intervention, then it seems likely that there might be, in the future, corresponding positive changes in the Science, Mathematics and English scores in the Oregon statewide subject tests for these students.

Given the high correlation between the BLOT and the achievement scores of individual students, it is surprising that some overall effects were so weak (e.g. Cohort C for Mathematics, and all cohorts for Reading & Literature). It is very difficult to account for this apparent contradiction. A correlation of 0.7 represents 49% of the variance, which would normally bode for strong effects in associated phenomena.

The CASE research in the UK showed that gains in cognitive development translated into improved student achievement some three years after the end of the intervention in national examination results in Science, Mathematics and English. Effect sizes between 0.3 and 1.0 standard deviations were reported (Adey & Shayer, 2002). It could be very interesting to see whether similar evidence of a far transfer effect can be detected in the student achievement data of Cohort B, for example, when these students complete their secondary education in June 2005. In the light of these British results, perhaps it is still too early to judge the full effects of the STEP on the school achievement of the Molalla River students.

Implications of the study

Although the results of this study were not as strong as those reported by Adey and Shayer (1994), they provide some evidence that the CASE intervention can benefit a broad spectrum of students, even when it is delivered in a less than optimal context. When the results of the STEP students are viewed in the "warts and all" context of its field setting, it is surprising that any improvement was detected in cognitive development and student achievement.

The teachers who began the STEP with their classes in September 1999 received very little professional development or technical support during the intervention. Indeed, some teachers who joined the school after the first year had no formal preparation at all. Furthermore, is likely that teacher confidence would have been adversely affected by the inclusion in the curriculum of a series of especially prepared lesson sequences, particularly if the rationale behind the intervention was poorly understood. Given the context in which the STEP was delivered, it seems unlikely that its moderate success could be attributed to the delivery of structured professional development to these teachers.

Other differences between the STEP and the original CASE studies were that none of the three STEP cohorts experienced all of the 30 CASE activities, and that the delivery of the intervention occurred over three school years for Cohorts A and B, as compared with two years in the UK CASE programs (and for Cohort C). Comparison of the results for Cohort B with their peers who missed the STEP intervention in 8th grade suggested that missing a year of the STEP might have inhibited some students' cognitive development and mathematics achievement. These students who missed a year of STEP experienced only 12 CASE lessons, as compared with the 16 or 21 lessons delivered to their peers in Cohort B. However, statistically significant overall results for cognitive change over the control were found for both Cohort B and C, using results of students tested on all occasions, although Cohort C students received only 13 CASE activities to the 16 or 21 delivered to Cohort B.

These important results suggest that the specific content of any particular CASE lessons might be a less important factor in the effectiveness of the intervention than the form or structure of the intervention experience. They also raise the question of just how many CASE lessons are necessary to yield a positive effect, and moreover, whether there are any indispensable lessons.

CASE activities 1 to 5 (control of variables and the fair test) were delivered to all three Cohorts in 1999-2000, and were revisited each subsequent year. The STEP teachers felt that the concepts central to these lessons, such as the control of variables and the fair test, would assist students in improving their scientific inquiry skills. Activities 1 to 5 were the only CASE lessons common to the three STEP cohorts. Cohorts B and C (but not A) received CASE activities 23 to 30 that covered the more advanced concepts of formal models, compound variables, and equilibrium. Perhaps it is not a coincidence that the delivery of this group of more challenging lessons appears to be associated with the late-onset cognitive development found in Cohorts B and C, but not in A.

Females in Cohort B made statistically significant gains over the control sample for cognitive change and Reading & Literature achievement. These results endorse the use of the STEP as an intervention that could be used in the Molalla River School District, and indeed other similar school districts, to promote the achievement of females. Furthermore, given that no significant relationship was found between cognitive change and the level of parent education, nor starting ability, it seems that the STEP also shows potential as a program to target low achieving females from all socioeconomic backgrounds. In the British CASE research, it was the sample of boys aged 12+ at the commencement of the intervention that appeared to gain most benefit, whereas it was the girls of that average age in Cohort B who showed the most cognitive growth in Molalla. Perhaps cultural effects, particularly those in the school setting, account for this difference between the performance of girls in the UK and in the Molalla River School District.

The progress of the three cohorts of STEP students showed some significant down turns, some of which had effect sizes comparable with the gains that have been attributed to the intervention. In particular, students in Cohort B had significantly lower BLOT scores in 2001 than in 2002. This was seen to be more marked for boys than for girls, and was found across the cohort, irrespective of the teacher. A similar downtrend was noticeable for male students in Cohort C when they were also in 8th grade, whereas the data for males in the control profile showed an upward trend at that age.

It seems illogical to propose that the 8th grade STEP students became less cognitively developed at that point in their adolescence. However, it is possible that factors associated with the circumstances of the BLOT test might have inhibited the display of the cognitive development of these students. The two 8th grade science teachers, Nancy and George, who delivered the STEP intervention at the time when the male students' BLOT scores fell, gave negative responses in the STEP teacher survey. Their answers showed that these teachers did not enjoy teaching the CASE activities, and further indicted that they saw little value in the intervention. Perhaps the disposition of these teachers lead to the BLOT being inadequately administered to these 8th grade students.

A further significant downtrend was found in the Mathematics and Reading & Literature OSS of Cohort B in the last year of the study. It is likely that the state-mandated tests would have been administered under more stringent conditions than those applying to the BLOT, and, indeed, teachers claimed to be unable to offer any school-based explanation for this fall in achievement scores. The downtrend occurred during the first year of high school for Cohort B, a transition that can be uncomfortable for some students. However, no evidence was found of a similar downturn in achievement in other samples of students as they entered the high school.

It is possible that the aftermath of the terrorist attacks of 9/11/01 might have inhibited the achievement of these students. The emergency procedures now routinely rehearsed in schools in the USA, as well as the explicit and pessimistic media coverage, are constant reminders to children that they live in troubled times. Furthermore, school counsellors reported that their case loads noticeably increased (personal communication, May 2003). However, if the events of 9/11 were indeed a traumatic factor, then it would suggest that Cohort B might have been more sensitive to this event than the students in either Cohorts A or C.

Another hypothesis to account for the fall in Mathematics and Reading & Literature scores is that there might be a delayed effect of cognitive development on student achievement. Perhaps it is not just a coincidence that the troughs in mean Mathematics and Reading & Literature achievement occur during the year following the fall in mean cognitive development for Cohort B and males in Cohort C. Furthermore, the profiles of Arthur and Ron indicate that these trends can be seen in the data of at least these two children, and are not merely statistical artifacts.

If there is indeed a relationship between these STEP participants' cognitive development and their subsequent student achievement, then this could be evidence of a far reaching teacher-associated effect. Nancy delivered eight CASE activities to her 8th grade classes, whereas George taught only three, and their other colleagues refused to participate in the intervention. No significant difference was found between the cognitive development of Nancy and George's students in the last two years of the study, despite George's students receiving five less CASE lessons in 8th grade. This suggests that the downtrend in student achievement is unlikely to be associated with the number of CASE activities delivered. Rather, it appears that the professional attitude of the 8th grade science teachers might have been an important factor affecting the achievement of the students in Cohort B.

What is it about the CASE intervention that makes it so successful in raising standards of cognitive development and student achievement? To this author there are striking similarities between the CASE lessons in the Thinking Science INSET video (Adey, 1993) and the 1999 TIMSS video segment (NCES 2003) of 8th grade mathematics teaching in Japan, regarded as exemplary by Stigler *et al.* in their 1999 report. The common features include a clear focus on high level concepts, thinking and problem solving, and an emphasis on students working in groups to derive solutions and explain their thinking. These characteristics were not evident in the TIMSS video clip of an 8th grade US mathematics lesson, although American teachers professed to including them in their practice (Stigler *et al*, 1999).

These common features might explain how the STEP intervention enhanced the cognitive development and Mathematics achievement of some students in the Molalla River School District. Although the teachers did not receive much training or support, their delivery of the CASE strategies exposed students to high-level Piagetian concepts, a focus on thinking and problem solving, and to working in groups to derive solutions and explain their thinking. If these are indeed the keys to enhancing student achievement, the implications are clear: if teachers can be educated to change their teaching style to enhance their students' operational thinking capacity, then achievement gains are likely to follow.

Recommendations to the Molalla River School District

Several general recommendations can be made to the administrators of the Molalla River School District based on the results of this study. These include an endorsement of the continued use of STEP to target school achievement, given the benefit to students. In particular, the improvement in cognitive level and Mathematics achievement, shown by many of the 6th and 7th grade Molalla River students, provides a convincing argument for continuation of the STEP.

In light of the responses to the teacher questionnaire and the results of CASE research in the UK, it is strongly recommended that a long-term program of CASE professional development is established in the Molalla River School District. It is unlikely that the considerable benefits of CASE, as seen in the UK, will apply to the STEP if the teachers are not educated to see the relevance of the intervention. Consideration should also be given to the delivery of a two-year program of CASE activities, preferably commencing at 7th grade.

A review of the science curriculum in the Middle School is also recommended, in order to find the best fit for the CASE activities and to generally “prune” the curriculum of any redundant content in order to release more time for the STEP. Adaptation of the CASE materials by the STEP teachers to better serve an American audience is also recommended, given the teachers’ comments regarding the language used in the CASE worksheets.

It was also clear from the teacher survey that technical assistance for the STEP teachers should be made a priority in order to reduce the teacher stress associated with time constraints and inadequate supply of equipment. The results of this study could provide a basis for a grant application to fund a program of long-term, effective professional development for the teachers, as well as the employment of a laboratory assistant/teacher aide.

Significance of the research

While there have been several replications of the original CASE project in other countries, this is the only CASE project to use Bond's Logical Operations Test and Oregon State Scores as test instruments. This implies that a significant aspect of this research is that the same true interval scale was used for the measurement of student achievement and cognitive change, so that a meaningful assessment could be made of children's cognitive development and school achievement. To my knowledge, no such studies of child development using Rasch measures for both achievement and development have been published.

Another feature of this research which makes it very different from other fully-funded CASE projects, is that the STEP teachers received very little professional development or in-class support. The number of CASE activities delivered to each cohort (and sometimes within each cohort) varied according to the teacher's disposition. The results of this study therefore show what can be achieved when the CASE activities are delivered in a sub-optimal setting.

Some of aspects of the STEP project have not been reported previously in the CASE literature: for example, comparisons between the results for students who experienced only part, or the whole, of the available intervention. Likewise, to my knowledge, there have been no previous reports of the effect of teacher attitude on the results of a CASE intervention.

Although the benefits to Molalla River students were modest in comparison with those claimed by the authors of the original CASE projects in Britain, the results of this study provide some evidence that student achievement really is not immutable if teachers can be educated to change their teaching style to enhance their students' operational thinking capacity.

REFERENCES

Adams, R.J. & Koo, S.-T. (1993). *Quest: the interactive test analysis system.* Hawthorn, Victoria: Australian Council for Educational Research.

Adey, P. (1992). Gender differences in the application of intellectual processes. *Studia Psychologica, 34*, 225-245.

Adey, P. (1993). *King's - BP INSET pack*. Poole: BP Educational Service.

Adey, P., & Shayer, M. (1990). Accelerating the development of formal thinking in middle and high school students. *Journal of Research in Science Teaching, 27*, 267-285.

Adey, P.S., & Shayer, M. (1991). An exploration of the long term far-transfer effects following an extended intervention programme in the high school curriculum. *Cognition and Instruction*, *11*, 1-29.

Adey, P., & Shayer, M. (1994). *Really raising standards.* London: Routledge.

Adey, P. & Shayer, M. (2002). Cognitive acceleration comes of age. In Shayer, M. & Adey, P. (Eds), *Learning Intelligence*. Milton Keynes: Open University Press.

Adey, P.S., Shayer, M. & Yates, C. (1995). *Thinking science* (2nd Ed.). London: Nelson.

Adhami, M., Johnson, D. & Shayer, M. (1998). *Thinking maths. The programme for accelerated learning in mathematics*. Oxford: Heinemann.

American Psychological Association (1999). Report of the Task Force on Statistical Inference. Available at: http://www.apa.org/journals/amp/amp548594.html [August 11 2003].

American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: Author.

Andrich, D. (1988). *Rasch models for measurement*. Newbury Park, CA: Sage.

Andrich, D. & Styles, I. (1994). Psychometric evidence of intellectual growth spurts in early adolescence. *Journal of Early Adolescence, 14,* 3, 328-344.

Andersen, E. B. (1985). Estimating latent correlations between repeated testings. *Psychometrika, 50*, 3-16.

Ausubel, D. (1968). *Educational psychology: A cognitive view.* New York: Holt Rinehart and Winston.

Baldwin, A. L. (1967). *Theories of child development*. New York: Wiley.

Barton, P.E., & Coley, R. J. (1998). *Growth in school: Achievement gains from the fourth to eighth grade.* Princeton, NJ: Educational Testing Service.

Becker, L. (2003). Effect Size calculator. Available at http://www.uccs.edu/~lbecker/psy590/es.htm [17 May 2003].

Ben-Hur, M. (Ed.) (1994). *On Feuerstein's instrumental enrichment*. Palatine, IL: IRI/Skylight.

Binet, A., & Simon, T. (1980). *The development of intelligence in children*. Nashville, TN: Williams Printing Co. (Original work published 1905).

Bond, T.G. (1976). *The development, validation and use of a test to assess Piaget's formal stage of logical operations.* Unpublished thesis, James Cook University of North Queensland: Townsville, Australia.

Bond, T.G. (1989). An investigation of the scaling of Piagetian formal operations. In P. Adey (Ed.) *Adolescent development and school science* (334-341). Lewes, UK: Falmer Press.

Bond, T.G. (1990). *An investigation of the structure of the Piagetian formal operations stage: An application of Rasch Analysis.* Unpublished thesis, James Cook University of North Queensland: Townsville, Australia.

Bond, T. G. (1994, June). Piaget's *Méthode Clinique*: There's more to it than we've been told. Paper presented at the Annual Symposium of the Jean Piaget Society, Chicago, USA.

Bond, T.G. (1995a). Piaget and measurement II: Empirical validation of the Piagetian model. *Archives de Psychologie, 63*, 155-185.

Bond, T.G. (1995b). Piaget and measurement I: The twain really do meet. *Archives de Psychologie, 63*, 71-87.

Bond, T.G. (1997). Measuring development. Examples from Piaget's theory. In L. Smith, J. Dockrell & P. Tomlinson (Eds.) *Piaget, Vygotsky and beyond.* London: Routledge.

Bond, T.G. (1998). Fifty years of formal operational research: the empirical evidence. *Archives de Psychology, 66*, 217-234.

Bond, T. G (2001). Building a theory of formal operational thinking: Inhelder's psychology meets Piaget's epistemology. In Tryphon, A. & Vonèche, J. (Eds.) *Working with Piaget: Essays in honour of Bärbel Inhelder* (65-83). Hove, Sussex: Psychology Press.

Bond, T.G. (2003). Relationships between cognitive development and school achievement: a Rasch measurement approach. In R. F. Waugh (Ed.), *On the Forefront of Educational Psychology*. New York: Nova Science Publishers (37-46).

Bond, T.G., & Fox, C.M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Erlbaum.

Bracey, G. W. (1987). Why the schools can't improve: the upper limit hypothesis. *Journal of Instructional Development*, *10*, 7-12.

Branson, R. K. (1998). Teaching-centered schooling has reached its upper limit: It doesn't get any better than this. *Current Directions in Psychological Science, 7*(4), 126-135.

Bunting, E. (1993). *A qualitative and quantitative analysis of Piaget's control of variables scheme.* Unpublished thesis, James Cook University of North Queensland: Townsville, Australia.

Campbell, J., Reese, C., & Dossey, J. (1996). *NAEP 1994: Trends in academic progress*. Washington DC: National Center for Educational Statistics.

Campione, J. C., & Brown, A. L. (1974). Transfer of training: effects of successive pretraining of components in a dimension-abstracted oddity task. *Journal of Experimental Child Psychology 18*, 398-411.

Campione, J. C., & Brown, A. L. (1978). Toward a theory of intelligence: contributions from research with retarded children. *Intelligence 2*, 279-304.

Carr, M., Borkowski, J. G. & Maxwell, S. E. (1991). Motivational components of underachievement. *Developmental Psychology, 27 (1)*, 108-118.

Carter, G., & Jones, M.G. (1994). Verbal and non-verbal behavior of ability-grouped dyads. *Journal of Research in Science Teaching, 31*, 603-619.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: a critical experiment. *Journal of Educational Psychology, 54*: 1-22.

Christiansson, D.J. (1983). *An investigation of the relationship between cognitive developmental stage and quantitative skills in college students.* Unpublished thesis, University of the South Pacific, Fiji.

Clackamas County (2003). Molalla Community Profile. Available at: www.co.clackamas.or.us [July 25 2003].

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd Ed.). Hillsdale, NJ: Erlbaum .

Coleman, J. S., & Associates (1966). *Equality of educational opportunity*. Washington, DC: US Department of Heath, Education and Welfare.

De Vries, R. (1974). Relationships among Piagetian, IQ and achievement assessments. *Child Development, 45*, 746-756.

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95(2)*, 256-273.

Embretson, S. E. (1991). A multidimensional latent trait model for measuring learning and change. *Psychometrika, 56*, 495-515.

Endler, L. C. (1998a). *Cognitive Development in a Secondary Science Setting*. Unpublished thesis, James Cook University, Townsville, Australia.

Endler, L. C. (1998b). Cognitive Development in a Secondary Science Setting. *The Queensland Science Teacher*, 25(2), 2-5.

Endler, L. C., & Bond, T. G. (2001). Cognitive development in a secondary science setting. *Research in Science Education*, 30(4), 403-416.

Endler, L. C., & Bond, T. G. (2002, June). Measuring changed minds: Empirical evidence from *Thinking Science* interventions in Australia and USA. Paper presented at the 10th International Thinking Conference, Harrogate, UK.

Epstein, H. T. (1986). Stages in human brain development. *Developmental Brain Research,* 30, 114-119.

Feuerstein, R., Rand, Y., Hoffman M., & Miller, M. (1980). *Instrumental enrichment: An intervention programme for cognitive modifiability.* Baltimore, MD: University Park Press.

Fischer, G. H. (1976). Some probabalistic models for measuring change. In D. N. M. de Gruijter & L. J. T. van der Kamp (Eds.), *Advances in psychological and educational measurement* (pp. 97-110). New York: John Wiley & Sons.

Fischer G.H., & Molenaar, I.W. (Eds.). (1995). *Rasch models: foundations, recent developments, and applications*. New York: Springer-Verlag.

Forman, E., & Cazen, C. (1985). Exploring Vygotskian perspectives in education: the cognitive value of peer interaction. In J. Wertsch (Ed.), *Culture communication and cognition: Vygotskian perspectives* (p. 323-347). Cambridge UK: University Press.

Gagné, R. M. (1965). *The conditions of learning*. New York: Holt Rinehart and Winston.

Ginsburg, H. P., & Opper, S. (1988). *Piaget's theory of intellectual development*. Englewood, NJ: Prentice Hall.

Gipps, C., & Murphy, P. (1994). *A fair test? Assessment, achievement and equity.* Buckingham: Open University Press.

Guilford, J. P. (1967). *The nature of human intelligence*. New York: McGraw-Hill.

Hautamäki, J. (1989). The application of a Rasch model on Piagetian measure of stages of thinking. In P. Adey (Ed.) *Adolescent development and school science* (342-350). Lewes, UK: Falmer Press.

Hautamäki, J., Arinen, P., Bergholm, B., Hautamäki, A., Kupiainen, S., Kuusela, J., Lehto, J., Niemivirta, M. & Scheinin, P. (1999). *The learning-to-Learn: The 6th grade.* National Evaluation Results. Helsinki: National Board of Education. (In Finnish).

Hayward, H. C., Brooks, P., & Burns, S. (1986). Stimulating cognitive development at developmental level: A tested non-remedial preschool curriculum for preschoolers and older retarded children. In M. Schwebel & A. Maher *Facilitating cognitive development: principles, practices, and programs* (pp. 127-147). New York: Haworth Press.

Holbech, J. D. & Thomsen, P. V. (2002). *Higher order thinking in Physics education (Hot Physics)*. Unpublished paper.

Hudspeth, W. J. & Pribram, K. H. (1990). Stages of brain and cognitive maturation. *Journal of Educational Psychology*, 2, 861-864.

Ingebo, G. S. (1997). *Probability in the measure of achievement*. Chicago: MESA.

Inhelder, B. (1954). Les attitudes expérimentales de l'enfant et de l'adolescent. *Bulletin de Psychologie,* 7 (5), 272-282.

Inhelder, B. (1963). Afterword: *The diagnosis of reasoning in the mentally retarded*. New York: John Day.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence: An essay on the construction of formal operational structures.* London: Routledge & Kegan Paul.

Iqbal, H. M., & Shayer, M. (1995). Distribution of Piagetian cognitive levels in middle school students and its match with science curricula. *Bulletin of Educational Research, XVII-XVIII* (1, 2).

Iqbal, H. M., & Shayer, M. (2000). Accelerating the development of formal thinking in Pakistan secondary school pupils: achievement effects and professional development issues. *Journal of Research in Science Teaching, 37(3)*, 259-274.

Jackson, S. (1965). The growth of logical thinking in normal and subnormal children. *British Journal of Educational Psychology, 35*, 255-258.

Johnson, J. S. & Newport, E. L. (1993). Critical periods in second language learning: the influence of maturational state on the acquisition of English as a second language. In M. H. Johnson (Ed.) *Brain development and cognition*. Oxford: Blackwell.

Jones, M. G., & Carter, G. (1994). Relationship between ability-paired interactions and the development of fifth-graders concepts of balance. *Journal of Research in Science Teaching, 31*, 847-856.

Jones, M., & Gott, R. (1998). Cognitive acceleration through science education: alternative perspectives. *International Journal of Science Education, 20(7)*, 755-768.

King, J., & Bond, T. G. (1996). A Rasch analysis of a measure of computer anxiety. *Journal of Educational Computing Research, 14*, 49-65.

Klauer, J. K. (1989). Teaching for analogical transfer as a means of improving problem-solving, thinking and learning. *Instructional Science 18*, 179-192.

Klein, P. S. & Aloni, S. (1993). Immediate and sustained effects of maternal mediating behaviours on young children. *Journal of Early Intervention, 17*, 1-17.

Kozulin, A. (2003). Psychological tools and mediated learning. In Kozulin, A., Gindis, B., Ageyev, V. S. & Miller, S. M. (Eds.) *Vygotsky's educational theory in cultural context* (pp. 15-38). Cambridge: Cambridge University Press.

Kozulin, A. (2000). Diversity of instrumental enrichment applications. In A. Kazulin and Y. Rand (Eds.) *Experience of mediated learning* (pp. 257-273). Oxford: Pergamon.

Kozulin, A. (1998). *Psychological tools*. Cambridge, Mass.: Harvard University Press.

Kozulin, A. (1990). *Vygotsky's Psychology: a Biography of Ideas*. Hemel Hempstead: Harvester Wheatsheaf.

Krogh, L. & Thomsen, P. V. (2001). Teaching style and learning outcomes. *Proceedings of GIREP-conference: Physics Teacher Education Beyond 2000*, 257-260.

Kuhn, D., & Angelev, J. (1976). An experimental study of the development of formal operational thought. *Child Development, 47*, 607-706.

Kuusela. J. & Hautamäki, J. (2002, June). *CASE and CAME came to Finland to enhance cognitive development: good results and surprising outcomes.* Paper presented at the 10th International Thinking Conference: Harrogate, UK.

Lake, D. C. (1996). *Quantitative approaches to the Piagetian analysis of temporal concepts held by Papua New Guinean trainee teachers*. Unpublished thesis, James Cook University of North Queensland: Townsville, Australia.

Lawson, A. E., & Blake, A. J. D. (1976). Concrete and formal thinking abilities in high school biology students as measured by three separate instruments. *Journal of Research in Science Teaching, 13 (3)*, 227-235.

Lawson, A. E., Karplus, R., & Adi, H. (1978). The acquisition of propositional logic and formal operational schemata during the secondary school years. *Journal of Research in Science Teaching, 15* (6), 465-478.

Lawson, A. E., & Snitgen, D. A. (1982). Teaching formal reasoning in a college biology course for preservice teachers. *Journal of Research in Science Teaching, 19*, 233-248.

Lawson, A. E., & Wollman, W. T. (1976). Encouraging the transition from concrete to formal cognitive functioning- an experiment. *Journal of Research in Science Teaching, 13(5)*, 413-430.

Leo, E. L., & Galloway, D. (1996). Conceptual links between Cognitive Acceleration through Science Education and motivational style: a critique of Adey and Shayer. *International Journal of Science Education*, 18(1), 35-49.

Lovell, K. (1961). A follow up of Inhelder and Piaget's *The Growth of Logical Thinking*. *British Journal of Educational Psychology, 52(2)*, 143-153.

Lovell, K., and Shields, J. B. (1967). Some aspects of a study of the gifted child. *British Journal of Educational Psychology, 37*, 201-208.

Macdonald, R. (1994). *"Why can't they understand this?" Is school science achievement constrained by cognitive development?* Unpublished MEd project, James Cook University, Townsville, Australia.

Mbano, N. (2003). The effects of a Cognitive Acceleration Intervention Programme on the performance of secondary school pupils in Malawi. *International Journal of Science Education*, 25, 71 -87.

McLellan, R., & Adey, P. (1999). *Motivational style, commitment, and cognitive acceleration: Is it only good schools which opt into "successful projects"?* Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.

McNamara, T. F. (1996). *Measuring second language performance.* Harlow: Addison Wesley Longman.

Michell, J. (1999). *Measurement in Psychology: critical history of a methodological concept.* Cambridge, UK: Cambridge University Press.

Mullis, I., Martin, M., Beaton, A., Gonzales, E., Kelly, D., & Smith, T. (1998). *Mathematics and science achievement in the final year of secondary school: IEA's Third International Mathematics and Science Study (TIMSS).* Boston: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.

Mullis, I. V. S., Martin, M. O., Gonzales, E. J., & Kennedy, A.M. (2003). *PIRLS 2001 International report: IEA's study of reading literacy achievement in primary schools*. Chestnut Hill, MA: Boston College.

National Center for Educational Statistics (2002). Program for International Student Assessment (PISA). Available at: www.nces.ed.gov/surveys/PISA [1 October 2002].

National Center for Educational Statistics (2003). Trends in International Mathematics and Science Study. Available at www.nces.ed.gov/timss [July 25 2003].

National Foundation for Educational Research (1979). *Science reasoning tasks*. Windsor: 16 Fen End, Over, Cambridge CB4 5NE, UK.

Newman, D., Griffin, P., & Cole, M. (1989). *The construction zone*. Cambridge, UK: Cambridge University Press.

Organization for Economic and Cooperative Development (2003). Program for International Student Assessment (PISA) data. Available at http://www.oecd.org/statistics [6 August 2003].

Oregon Department of Education (2001). Oregon assessment system. Available at http://www.ode.state.or.us [11April 2001].

Paour, J.-L. (1992). Induction of logic structures in the mentally retarded: an assessment and intervention instrument. In H.C. Haywood and D. Tazuriel (Eds.), *Interactive assessment*. New York: Springer-Verlag.

Peaker, G. (1975). *An empirical study of education in twenty-one countries: A technical report.* New York: Wiley.

Piaget, J. (1953). *Logic and Psychology*. Manchester: The University Press.

Piaget, J. (1972). Intellectual evolution from adolescence to adulthood. *Human Development, 15*, 1-12.

Rasch, G. (1960/80). *Probabalistic models for some intelligence and attainment tests.* Chicago: University of Chicago Press.

Rogoff, B. (1995). Observing sociocultural activity on three planes: Participatory approbation, guided participation and apprenticeship. In J. Wertsch, P. Del Rio, & A. Alvarez (Eds.), *Sociocultural studies of the mind* (pp. 139-164). Cambridge, UK: Cambridge University Press.

Rosenthal, D. A. (1979). The acquisition of formal operations - the effect of two training procedures. *Journal of Genetic Psychology, 134*, 125-140.

Rousseau, J-J. (1762). *Emile*. Paris: Duchesne. English edition, 1911. London: Dent.

Schmidt, W. H., Jorde, D, Cogan, L., Barrier, E., Gonzalo, I., Moser, U., Shimizu, K., Valverde, G., McKnight, C., Prawat, R., Wiley, R., Raizen, S., Britton, T., & Wolfe, R. (1996). *Characterizing pedagogical flow: An investigation of mathematics and science teaching in six countries*. Dordrecht: Kluwer.

Schmidt, W. H., McKnight, C., Cogan, L.G., Jakwerth, P., & Houang, R. (1999). *Facing the consequences: Using TIMSS for a closer look at U.S. mathematics and science education.* Dordrecht: Kluwer.

Shayer, M. (1993). King's College CASE-INSET 1991-1993: Effect on cognitive development. Press report available from School of Education, King's College, University of London.

Shayer, M. (1999). Cognitive acceleration through science education II: its effects and scope. *International Journal of Science Education, 21(8)*, 883-902.

Shayer, M. (2002). Not just Piaget, not just Vygotsky, and certainly not Vygotsky as an alternative to Piaget. In Shayer, M. & Adey, P. (Eds), *Learning Intelligence* (pp. 179-195). Milton Keynes: Open University Press.

Shayer, M., & Adey, P. (1981). *Towards a science of science teaching*. London: Heinemann.

Shayer, M., & Adey, P. S. (1992a). Accelerating the development of formal thinking in middle and high school students: III. Testing the permanency of effects. *Journal of Research in Science Teaching, 29*, 1101-1115.

Shayer, M., & Adey, P. S. (1992b). Accelerating the development of formal thinking in middle and high school students: III. Testing the permanency of effects.

Journal of Research in Science Teaching, 29, 1101-1115.

Shayer, M., & Adey, P. S. (1993). Accelerating the development of formal thinking in middle and high school students: II. Post project effects on science achievement. *Journal of Research in Science Teaching, 29*, 81-92.

Shayer, M., & Adey, P. (2002). *Learning Intelligence.* Buckingham, Philadelphia: Open University Press.

Shayer, M., & Beasley, F. (1987). Does Instrumental Enrichment work? *British Educational Research Journal, 13(2)*, 101-119.

Shayer, M., Küchemann, D. E., & Wylam, H. (1976). The distribution of Piagetian stages of thinking in British middle and secondary school children. *British Journal of Educational Psychology,* 46, 164-173.

Shayer, M., & Wylam, H. (1978). The distribution of Piagetian stages of thinking in British middle and secondary school children. II - 14- to 16- year olds and sex differentials. *British Journal of Educational Psychology, 48*, 62-70.

Siegler, R., Liebert, D., & Liebert, R. (1973). Inhelder and Piaget's pendulum problem: Teaching preadolescents to act as scientists. *Developmental Psychology, 9, 1*, 97-101.

Skinner, B. F. (1976). *About behaviourism*. New York: Vintage Books.

Smith, L., & Knight, P. (1992). Adolescent reasoning tests with history content. *Archives de Psychologie, 60*, 225-242.

Sokal, R. R. & Rohlf, F. J. (1995). *Biometry, the principles and practice of statistics in biological research* (3rd Ed.). New York: W. H. Freedman & Co.

Stanbridge, B.A. (2002). A radical constructivist approach to high school science teaching. Unpublished thesis, James Cook University: Townsville, Australia.

Stigler, J.W., Gonzales, P., Kawanaka, T., Knoll, S. & Serrano, A. (1999). *The TIMSS* *videotape classroom study: methods and findings from an exploratory research project on eighth-grade mathematics instruction in Germany, Japan and the United States.* Washington D.C.: US Dept of Education, Office of Educational Research and Improvement, NCES.

Styles, I. (1999). The study of intelligence - the interplay between theory and measurement. In Anderson, M. (Ed.) *The development of intelligence* (pp. 19-42). Hove, UK: Psychology Press.

Suter, L. E. (2000). Is student achievement immutable? Evidence from international studies on schooling and student achievement. *Review of Educational Research, 70(4)*, 529-545.

Tanner, J. M. (1978). *Foetus into man: Physical growth from conception to maturity.* London: Open Books.

Thurstone, L. L. (1969). *Primary mental abilities*. Chicago: University of Chicago Press.

Tzuriel, D. (1997). The relation between parent-child MLE interactions and children's cognitive modifiability. In A. Kozulin (Ed.), *The ontogeny of cognitive modifiability* (pp. 157-180). Jerusalem: ICELP.

Tzuriel, D. & Feuerstein, R. (1992). Dynamic group assessment for prescriptive teaching: differential effects of treatments. In H.C. Haywood and D. Tazuriel (Eds.), *Interactive assessment*. New York: Springer-Verlag.

Tzuriel, D., Kaniel, S., Kanner, E., & haywood, H. C. (1997). The effects of the "Bright Start" program in kindergarteners on transfer and academic achievement. *Early Childhood Research Quarterly 14 (1)*, 111-141.

US Census Bureau (2002). Molalla demographic data. Available at www.census.gov [October 2 2002].

von Aufschnaiter, C. (2003). Theoretical framework and empirical evidence on students' cognitive processes in three dimensions of content, complexity, and time. *Journal of Research in Science Teaching*. In press.

Vygotsky, L. S. (1978). *Mind in society*. Cambridge, Mass.: Harvard University Press.

Vygotsky, L. S. (1981). The genesis of higher mental functions. In J. Wertch (Ed.), *The Concept of Activity in Soviet Psychology.* Armonk, NY.: M. E. Sharpe.

Wertsch, J. V. (1985). *Vygotsky and the Social Formation of Mind.* Cambridge, Mass.; London, England: Harvard University Press.

Wright, B. D. (1979). *Best test design.* Chicago: MESA Press.

Wright, B. D. (1985). Rasch measurement models. In Husen, T. and Postlethwaite, T. N. (Eds.), *The International Encyclopaedia of Education,* Oxford: Pergamon.

Wright, B.D.(1998). Estimating measures for extreme scores (Online). Available at http://mesa.spc.uchiago.edu/rmt/rmt122h.htm (October, 1998).

Appendix 1

Case activity 27:

Floating and Sinking

THIS APPENDIX HAS BEEN REMOVED DUE TO
COPYRIGHT RESTRICTIONS

Appendix 2

STEP Teacher Questionnaire

THIS APPENDIX HAS BEEN REMOVED DUE TO COPYRIGHT RESTRICTIONS