

EVALUATION REPORT OF THE NEW BASICS RESEARCH PROGRAM

John Ainley



Australian Council for Educational Research

July 2004

This work was funded by the Queensland
Department of Education and the Arts.

©The State of Queensland
(Department of Education and the Arts) 2004

CONTENTS

Tables.....	iii
Figures	iii
Executive Summary.....	v
1 Introduction.....	1
Background.....	1
Key Elements.....	1
The Trial.....	2
Research and Evaluation.....	3
Scope.....	3
Organisation of the Report.....	3
2 Aspects of Methodology	4
Purposes.....	4
Methods.....	4
Mixed Methods.....	4
Formative Evaluation.....	5
Summative Evaluation.....	7
3 Research Framework	9
Research Questions.....	9
Major themes from the research	9
Trial Schools and Comparison Schools.....	10
Samples and Precision	10
Data Sources	11
Summary.....	15
4 Rich Tasks and Student Work.....	16
The Concept of Richness.....	19
Construct Validity: Recognition of Richness	19
Comparisons of Rich Tasks and Other Samples of Student Work.....	21
Teachers' Responses to Implementing Rich Tasks	23
Samples of Student Work	25
Impact on School Work	26
Summary.....	27
5 Teaching and Learning.....	29
The Classroom Observation Study.....	30
The Enacted Pedagogy Study.....	34
Summary.....	36

CONTENTS (Cont)

6 Assessment.....	37
Teachers' Conceptions of Assessment	37
Teacher Assessments of Rich Tasks.....	38
The Rich Task Assessment System.....	39
Profiles of Results on Rich Task Assessment in 2003	41
Summary.....	43
7 Student Achievement.....	44
The International Schools Assessment	44
Statewide Literacy and Numeracy Assessments.....	48
World Class Tests.....	49
Summary.....	50
8 School Organisation and Structure.....	52
The Teacher Surveys	52
The Thirteen School Study	54
Community Linkages.....	57
The Role of the Critical Friend	57
Summary.....	58
9 Understandings, Responses and Other Issues.....	60
School Satisfaction	60
Knowledge of New Basics.....	62
New Basics Discussion Lists.....	63
Responses to New Basics	63
Costs and Feasibility.....	64
Summary.....	64
10 Summary and Conclusions.....	66
The Research Program.....	66
Summary Results Concerning Impact	67
Strengths and Weaknesses	69
Viability and Scope for Implementation.....	72
Areas for Review, Research and Development	73
In Conclusion.....	74
References.....	75

TABLES

Table 1	Project Data Sources in New Basics Research Program	12
Table 2	Summary of Rich Tasks for Years 1 to 3	16
Table 3	Summary of Rich Tasks for Years 4 to 6	17
Table 4	Summary of Rich Tasks for Years 7 to 9	18
Table 5	Rankings of Richness in Student Work.....	22
Table 6	Dimensions and elements of productive pedagogies	29
Table 7	Mean Scale values for Productive Pedagogies Dimensions	32
Table 8	Means and Standard Errors for EPA Dimension Scores by Year Level and Trial Status.....	35
Table 9	Classification of Level of Agreement in Grade Ratification	40
Table 10	Mean Scores on ISA Tests by Time, Year Level and Trial Status.....	46
Table 11	Reduction of the Gap in Median Scores for Trial and Non-Trial Schools between 2001 and 2003	49
Table 12	Effect Size Measures for Intended Impact of New Basics	68

FIGURES

Figure 1	Dimensions and Elements of Richness in New Basics	20
----------	---	----

EXECUTIVE SUMMARY

New Basics was an initiative by Education Queensland that sought to provide a framework for curriculum, pedagogy and assessment in Queensland schools. The framework defined areas of learning, approaches to teaching, modes of assessment and standards for student development. *New Basics* was manifest in three key elements: curriculum, pedagogy and Rich Tasks. Curriculum was formulated around four curriculum organisers (life pathways and social futures; multi-literacies and communication; active citizenship; and environments and technologies) rather than eight Key Learning Areas. Pedagogy referred to teaching and learning practices based on intellectual quality, connectedness, supportive classroom environments and recognition of difference. Rich tasks were a form of authentic assessment intended to provide the basis for assessing and reporting learning outcomes.

The Trial

Thirty-eight schools began with *New Basics* in 2000 and started implementing the framework in 2001 with the trial being completed in 2003. There was a research program associated with the *New Basics* trial conducted by the Assessment and *New Basics* Branch of Education Queensland with advice from the *Framework Research Advisory Group*. The focus of the research program was on whether *New Basics* resulted in desirable changes, whether those changes were likely to be accepted and whether it was feasible to extend the program beyond the trial.

The present external evaluation reviewed data and research findings, research methodologies and provided an evaluation report at the end of the trial. This external evaluation was based on the research reports from the research program and was not a large-scale study that utilised its own data. The Australian Council for Educational Research (ACER) was contracted by the Queensland Department of Education and the Arts to conduct this external evaluation.

The Research Program

The most important sources of data that were used in the research program were:

- Samples of the work of students on Rich Tasks and from everyday school activities.
- Results of observations of teachers in classrooms and a survey of students regarding their classroom experience.
- Records of teacher assessments of Rich Tasks and the processes of the moderated school-based assessment system.
- Measures of student achievement on three different types of achievement tests.
- Information about school organisation and structure derived from a qualitative study in 13 schools.
- A range of surveys concerned with responses to the *New Basics* and specific aspects of it.
- Records routinely gathered by the school system as part of its management processes.

Results

Rich Tasks and Student Work

It was found that Rich Tasks in 2003 from Year 6 students in trial schools exhibited higher levels of richness than work from high achieving Year 6 students in other schools. That difference was not as evident at Year 3 and Year 9. In Year 3 the tendency was towards lower levels of richness in trial schools and in Year 9 the levels of richness was the same in trial and non-trial schools. In terms of other student work there were differences in the intellectual quality of work in 2003 from trial schools and other schools at Years 8 and 9, in the direction of higher intellectual quality work from the trial schools. The difference appeared to be associated with the extent to which work was connected to the outside world.

Teaching and Learning in Classrooms

Observations of classroom teaching and learning practices suggested that there was a tendency, although not statistically significant, towards higher scores on intellectual quality, supportive classroom environment and recognition of difference dimensions in trial compared to non-trial schools. The small number of teachers involved limits the extent that this can be regarded with certainty. Student responses to a questionnaire about their classrooms indicated differences between trial and comparison schools on connectedness, classroom environment and recognition of difference among Year 5 classes such that more positive responses were recorded in trial schools.

Student Achievement

There was no difference between trial and comparison schools on the International School Assessments. More detailed analysis indicated that there was no general impact of being in the trial on relationships of achievement with socioeconomic background or context, or on the difference between Indigenous and non-Indigenous students (but there was an effect for Indigenous students reading in Year 6). There was no evidence of a decline in performance on the statewide literacy and numeracy tests and possibly a small reduction in the gap between trial schools and all other schools on those tests.

School Satisfaction

On the measures of school satisfaction in 2003 there were higher scores on satisfaction with quality curriculum and learning outcomes for primary students in trial schools compared to non-trial schools. This difference had not been evident in 2001 or 2002 and was not evident for secondary school students in 2003.

Strengths and Weaknesses

Quality of Student Work

A major strength of the New Basics program was on the nature and quality of student work. Providing the opportunity, incentive and structure for student work of high quality represents an important strength of the trial. In Year 6, and probably more generally in the upper primary years, there seemed to be greater opportunity to implement and develop Rich Tasks than at the other two juncture years.

Development of an Assessment System

Another strength of the New Basics was the establishment of an assessment system during the trial. This was central to establishing standards for student work and the implementation of the Rich Tasks. It is notable that richness became most evident only when the moderated assessment process was imminent. The introduction of this system provided a vehicle for reporting on the quality of student work. The moderated assessment system appeared to be crucial to the implementation of the Rich Tasks in the way that they were intended.

Changes in Approaches to Teaching

The evidence regarding changes in teaching and learning was more mixed. Changes became evident in the third year of the trial but it is harder to establish that the changes were greater in trial than non-trial schools. The New Basics emphasised wider use of certain teaching strategies and there is some evidence of that having taken place.

Achievement

Measures of achievement that go beyond samples of student work are important to determine the extent to which learning can be applied in new situations and to new problems. There was no impact of being in the New Basics trial on student achievement on the three assessment measures. New Basics literature did not argue that there would be an impact on literacy and numeracy skills but it did argue that it would promote higher order thinking skills. The International Schools Assessment (ISA) sought to measure higher order thinking in the context of reading literacy, mathematical literacy and writing and did not show a benefit from participation in the trial. There remains a need to articulate better the higher-order thinking that is envisaged by New Basics and to establish indicators that go beyond the work that students do.

Congruence with Other Aspects of the School System and its Context

The study identified that changes involving the roles of teaching staff, teacher release for planning and the provision of spaces for teachers to work together were required to implement New Basics. Changes were also needed in matters such as space utilisation, timetabling, subject availability and budgeting. The movement of students between schools appeared to be an issue that affected some students being able to complete Rich Tasks. The extension of the Year 9 Rich Tasks back to Year 7 caused problems where high schools drew students from multiple primary schools. There appeared to be a need to see how the New Basics organisers and the traditional disciplines could be reconciled theoretically and organisationally.

Differential Impact

From several perspectives there was evidence that the effects of New Basics differed between Year levels. The impact of New Basics appeared to be greatest in the upper primary years. In lower secondary school there was mixed evidence that suggested that with adaptation the Rich Tasks could be implemented with benefit. In the lower primary school the Rich Tasks appeared to have less impact than at the other two levels. There is not much evidence from the trial about equity aspects of New Basics and the little evidence that was available indicated that the relationship between achievement and

socioeconomic status was the same in trial and non-trial schools. In general there was no difference between trial and non-trial schools in the relationship between Indigenous and non-Indigenous status except for a higher level of growth in reading for Indigenous (compared to non-Indigenous) students in Year 6.

Viability and Scope for Implementation

Almost all of the trial schools agreed to continue to operate New Basics beyond the period of the trial. Of course their commitment to continue does not necessarily indicate that New Basics would be adopted in schools that had not sought to participate in the trial. A moderated assessment system linked to the Rich Tasks is important and would be one of the larger ongoing costs of New Basics. The moderated assessment system appeared to be a crucial element of support to the implementation of the Rich Tasks in the way that they were intended. Given that the trial operated most effectively in the upper primary school the most fruitful area for application of New Basics and New Basics principles could be the upper primary school. Implementation in high schools would probably require a review of the span of three years and possibly stronger links to the major disciplines of the curriculum. It is not clear from the trial whether the full span of three years and mapping curriculum is necessary to achieving all the goals associated with New Basics. A framework that focussed on tasks extending over a full year might fall short of the ideals of New Basics but might be more readily implemented within schools where there is a high degree of mobility of students and teachers.

Areas for Review, Research and Development

The New Basics trial provided the opportunity to examine a wide range of issues. The research program associated with the trial investigated a great many of those issues. One of the consequences of any program of research is that it identifies further issues for investigation and highlights aspects of the program that were not previously identified as needing exploration. There is a case for reviewing the evidence that has been accumulated and to use that evidence to guide the adaptation of Rich Tasks to environments where implementation proved difficult. It would be useful to replicate the teaching strategies study on a wider basis so that data on teaching could be linked to student outcome measures that captured change over a given period of time. New Basics literature discusses the Rich Tasks as being designed to engage students. It would be useful to investigate the extent to which students become engaged in those tasks and how they become engaged. Analysis of the Rich Task assessments conducted with data from the moderation process indicated that there was a common element that accounted for about half the variation in grades in addition to features unique to each Rich Task. It is important to understand better the nature of that common element and of the clusters among the unique elements identified in the research report.

INTRODUCTION

Background

New Basics was an initiative by Education Queensland that sought to provide a framework for curriculum, pedagogy and assessment in Queensland schools¹. The framework defined areas of learning, approaches to teaching, modes of assessment and standards for student development. It was intended to improve the learning of students in a variety of domains and was linked to a strategic plan for Education Queensland (*Queensland State Education 2010*) that stressed the importance of improved teaching and learning as well as flexible, diverse and effective school programs. *New Basics* provided a manifestation of the strategic plan expressed in terms of curriculum, teaching, learning and assessment. To a large extent *New Basics* was grounded in the perspectives and inferences drawn from the Queensland School Reform Longitudinal Study (QSRLS). QSRLS was a three-year study of classroom pedagogy and student outcomes that began in 1998 and was one of several research studies in the late 1990s that mapped the state of teaching and learning in Queensland schools and sought to identify ways in which improvements might be nurtured (University of Queensland, 2001).

New Basics was based on several inter-connected principles outlined in the *New Basics Technical Paper* (Education Queensland, 2000). Foremost was the principle that improved student outcomes requires a “systematic, principled and practical coordination of curriculum, teaching and assessment”. Those outcomes, it was argued, should be oriented to future workplaces, technologies and cultures and the changes in policies and practices intended to achieve those outcomes need to be grounded in research on school practice. It was also argued that improved outcomes required high levels of teacher professionalism, sustained intellectual work and shared ownership of reform. Finally, it was asserted that important, common learning should address the aspirations of the most at-risk and culturally diverse communities.

Key Elements

The New Basics Initiative was manifest in three key elements (called ‘conceptual pivots’): curriculum, pedagogy and Rich Tasks (Education Queensland, 2001 a).

Curriculum

As noted previously the term “new basics” was used to refer both to the overall framework (or design) and to organising elements of the curriculum (Education Queensland, 2001b). In the latter sense *New Basics* was intended as a means of organising curriculum content in schools in a way that reflected the contemporary context in, and for which, children learn. The curriculum organisers were: life pathways and

¹ The term “new basics” was previously used by the Ministerial Consultative Council on Curriculum (MCCC) in 1989 to refer to knowledge categories established as a result of consultations: communication – literacy, oracy, numeracy; citizenship in society – rights and responsibilities for effective participation in political, legal and economic systems; cultural and human environment; the individual in society; the natural and physical world and the pure sciences and new technologies (application and implications).

social futures; multi-literacies and communication; active citizenship; and environments and technologies. These organisers were argued to provide an alternative to the key learning areas that had been a feature of Australian education since the early 1990s. It was argued that the new basics were transdisciplinary but drew upon traditional disciplines (which are not the same as Key Learning Areas). However, it was also argued that *New Basics* was not just a matter of curriculum re-organisation.

Teaching and learning

The term “productive pedagogies” was used to describe classroom strategies. It was intended that teachers would develop strategies in relation to what they were teaching and the characteristics (backgrounds and learning styles) of their students. The pedagogies focused on intellectual quality, connectedness, supportive classroom environments and recognition of difference. There was a suite of 20 productive pedagogies including examples of practice. Productive Pedagogies also operated as a theme or focus in professional development on a broader basis outside the trial.

Rich tasks

Rich tasks were a form of authentic assessment intended to provide the basis for assessing and reporting learning outcomes. The Rich Tasks had a pedagogical as well as an assessment purpose and a large part of the rationale for them emphasised the role that they played in student learning. Notwithstanding their pedagogical role, Rich Tasks were a means through which students displayed their understandings, knowledge and skills and they provide the basis for the assessable and reportable outcomes of learning. There were three reporting points: at the end of Years 3, 6 and 9. A set of 20 Rich Tasks covered the three three-year spans. Each of the suites of Rich Tasks covered a balance of fields of knowledge and generic skills. This set of tasks had been published as a teacher booklet and distributed to trial schools. It formed the basis of how New Basics schools reported to parents and the system on student achievement. Rich Task assessment, and associated moderation, was envisaged as a key element of *New Basics* and as a major driver of changes in other elements of *New Basics* (Education Queensland, 2001 a: 4).

The Trial

A number of Queensland schools participated in a four-year trial of *New Basics*. In phase one of the trial, 38 schools began with New Basics in 2000 and started implementing the framework in 2001 with the trial being completed in 2003. A further 21 schools joined the project with preparatory activities in 2001 and implemented the framework in their classrooms from 2002 to 2004. According to the *New Basics Technical Paper* the aims of the trial were to: “fully develop and articulate the New Basics, Productive Pedagogies and Rich Tasks into mature and generalisable approaches to curriculum, pedagogy and assessment that lead to verifiable improvements in student outcomes”; and to “determine the requisite school organisational capacities needed to successfully implement these approaches” (Education Queensland, 2000: 106). The New Basics Research Program was intended to conduct studies to establish the extent to which these aims were being achieved. There was a developmental aspect to the trial that involved further conceptualisation and elaboration of the ideas in the framework as well as implementation of the framework in schools (starting with Years 1, 4 and 7) using professional development, organisation of resources and support.

Research and Evaluation

There was an extensive research program associated with the *New Basics* trial. The Assessment and New Basics Branch of Education Queensland conducted the *New Basics Research Program* with advice and guidance from the *Framework Research Advisory Group* (FRAG). FRAG was an independent expert group that guided the overall design of the research program, provided technical advice on methodology and instruments, assisted in the identification of patterns and trends in data and provided assistance in the production of the final report. In addition to the small group of staff in the Assessment and New Basics Branch, some research activities were contracted to outside analysts and researchers. The focus of the research program was on the viability of New Basics (Education Queensland, 2003a: 3) in terms of whether it resulted in desirable changes, whether those changes were likely to be accepted and whether it was feasible to extend the program beyond the trial.

This external evaluation was concerned with the merit of the program, specifically the part New Basics should play in future strategies to improve student learning. The role of the evaluation was to review data and research findings, comment on the conduct of the research in each year of the trial and provide an independent evaluation report at the end of the trial. Although the evaluator had contact with the researchers, the evaluation was at a distance from them in operations and reporting. The evaluation was based on the research reports from the research program but the conclusions drawn are independent. The evaluation was not a large-scale study that duplicated the research program and utilised its own data. It was based on the data gathered and analysed by the staff involved in the research program and the reports written by those staff. However, it not just based on the stated conclusions but also on an appraisal of the evidence that underpinned those conclusions.

Scope

The New Basics trial extended over three years and the research program produced 25 research reports that encompassed a range of issues concerned with New Basics. The trial and the research program thus represent a significant and complex endeavour. There are all too few studies that extend so far over time and whose scope is as wide as this.

Organisation of the Report

The summative evaluation report is organised around ten chapters. Following this introduction the second chapter provides an outline of some aspects of methodology and the third provides a description of the research framework for the New Basics Research Program. There are then six chapters concerned with the main themes of the evaluation: Rich Tasks and student work, teaching and learning, assessment, student achievement, school organisation and structure, and understandings and responses. The final chapter focuses on broad conclusions.

ASPECTS OF METHODOLOGY

Evaluation and research both use similar methods of empirical enquiry, procedures for gathering evidence and analysis of data. Both are concerned with discovering previously unknown knowledge. However, evaluation and research differ with regard to the generality of their findings, with evaluation focussing on information specific to a particular program, and with respect to purpose. Evaluation is a form of systematic educational enquiry that focuses on questions of the worth or merit of programs in a particular context (Scriven, 1967). For that reason it is important to be explicit about what is being evaluated. Within this general orientation evaluation can, and increasingly does, involve a wide variety of forms and incorporate a wide range of methods for gathering and analysing evidence (Kellaghan, Stufflebeam & Wingate, 2003). Evaluation can be concerned with the origins of a program, the process of implementation of a program as well as the impact of a program. Moreover evaluation can be concerned with formative purposes (e.g. deciding how to improve a program) or summative purposes (e.g. deciding whether a program should be continued).

Purposes

The present evaluation had both formative and summative purposes and was concerned with both the research program and the New Basics. In terms of its formative role the evaluation provided comment on the conduct of the research program during the trial and a progressive review of research findings. In its summative role at the conclusion of the trial it provided an appraisal of the extent to which the data and analysis provided a sound basis for the conclusions. Thus it considered potential threats to the internal (are the findings consistent with the data) and external (can the findings be extrapolated beyond the study) validity of the research. As part of its summative role the evaluation provided an overview, or meta-analysis (in the non-technical sense of that term), of the results of the research program and an interpretation of the findings in the wider sense of the inferences that relate to further developments in education in Queensland.

Methods

The predominant method used in the evaluation was the analysis of documents produced as part of the research program and the secondary analysis of data from the research program. This was supplemented by discussions with the research staff regarding the particular investigations that they were undertaking and with other staff regarding the framework within which the investigations were organised. Attendance at a number of meetings of the FRAG provided both insights into decisions about purposes and research design as well as an opportunity to contribute to deliberations about methodology. Only a very limited number of visits to schools (three) were undertaken to provide a context and not to gather data.

Mixed Methods

The overall research framework for the New Basics Research Program fitted a “mixed-methods” model (Tashakkori & Teddlie, 2003: Frechtling & Sharp, 1997). Mixed methods approaches to educational and social research seek to combine the strengths of

complementary approaches to research (sometimes crudely delineated as quantitative and qualitative). In the case of the New Basics Research mixed methods were appropriate for two reasons. Firstly, mixed methods were appropriate because they provided comprehensiveness (applying research methods that were most appropriate for particular issues). In other words different research questions in the framework required different methods of investigation. Secondly, mixed methods were appropriate because they provided a more robust basis for inference by combining findings from different approaches to the same issue.

Despite the benefits of mixed methods approaches there are difficulties in terms of the ways in which inferences are drawn from different types of study and in integrating the findings from different types of research (Tashakori & Teddlie, 2003). For example, techniques for the integration of findings from different studies have been typically based on quantitative research synthesis (eg. meta-analysis) or narrative forms of review. In integrating the findings from different elements in a mixed methods approach quantitative synthesis was not sufficient because it depended on numerical estimates of effect sizes as the basis for integration. For that reason mixed methods models usually combine narrative approaches with some estimation of the size of effects where that is possible.

Formative Evaluation

The formative evaluation process involved both formal reports and informal communication with researchers, the director of the Assessment and New Basics branch, and the FRAG. The formal reports had different emphases over time as the trial progressed. The first was concerned with a commentary on research plans. The second was a commentary on an organising framework and progress with analyses. The third provided an overview of coverage of data and research projects. The following are the main issues raised in the formative evaluation process.

Samples of Student Work and Moderation

The formative evaluation drew attention to the lack of use of student work samples in educational research and the issues of resources involved and defining dimensions for analysis (in this case richness). The definition of richness dimension is well established in the research reports. It raised issues of sampling of schools and work within schools in terms of numbers of students per school for stability and the basis of selection to ensure lack of bias. It discussed the issues that could arise from assessors knowing whether work was from a trial or non-trial school.

The formative evaluation commented on the desirability of examining between-task effects as well as between-school and between-marker effects in the moderation process. It suggested the use of data to provide annotated work samples that could be used to clarify expectations in relation to student work.

Teaching and learning

The formative evaluation commented on the importance of the time-series dimension in observations within trial and non-trial schools. It argued for the importance of a large sample of schools so as to be able to investigate the range within each but recognised the sample would be limited by resources. It suggested that there should be more than three

observations per class in any year. It noted that, although it would not be possible for observers not to be told whether the school was a trial school or not, this should be considered a potential source of bias in a high inference measure.

The formative evaluation suggested collecting data on the classrooms (at some year levels) using a learning environment inventory completed by students. It was noted scales similar to the dimensions of new basics but independent of the initiative itself would be useful. Later a survey using the Enacted Pedagogical Assessment (EPA), which was designed to assess student perceptions of the teaching in their classrooms, was conducted.

Following from the issue of the extent to which the New Basics reduced alienation in the middle years of schooling, it was suggested that measures of student engagement and interest be obtained through student surveys over time in trial and non-trial schools.

The formative evaluation canvassed whether it might also be valuable to gather additional data about what happens in classrooms using “teacher logs” similar in concept to those developed by the University of Michigan in its Study of Instructional Improvement. This would have been resource intensive in development and was deemed beyond the scope and resources of the New Basics Research Program.

Achievement Measures

The formative evaluation encouraged the analysis of statewide literacy and numeracy tests. It suggested the use of multilevel regression analyses to establish growth (or change) measures from Year 3 to Year 5 and from Year 5 to Year 7 over the course of the trial. It was argued that growth scores would be more sensitive to pedagogical effects than scores at any point in time. It noted the use of these methods to establish intake-adjusted school effects in New South Wales. This recommendation proved to be difficult to implement because it was not possible to establish enough matching identification across Year levels.

The formative evaluation noted the technical issues in the use of the World Class Tests and suggested seeking an alternative assessment that probed higher order skills because the issue was important to New Basics. It also commented that the sampling procedures (taking the top ten per cent of students in each school) would limit inferences from the data and suggested a different instrument and procedure.

The formative evaluation recommended the use of the International Schools Assessment (ISA), and the student background data to be used in its analysis, for modelling achievement growth. ISA was successfully conducted and analyses conducted as envisaged. It was argued that it was important to gather information about the background (that could be used to generate measures of socioeconomic status etc) of individual students so that equity effects can be investigated. Parental occupations, educational attainments and country of birth are typically considered the minimum for conducting analyses that provide indications of the social distribution of achievement. Information about Indigenous status was also important. The collection of individual-level data was necessary to conduct multi-level analysis at student and school level.

Other Matters

The formative evaluation recommended mapping the content of activities using a discipline-based map in a sample of trial schools compared with non-trial schools. It also recommended a framework that integrated two alternative representations being canvassed. In the process of developing the report structure a different framework was developed and used.

The formative evaluation indicated that any analysis of attendance needed to overcome the skewed nature of the data and pointed to the thorough analysis of attendance data reported for South Australia (see Rothman, 2002) that could be replicated in Queensland. However, attendance data in Queensland were not in a form that could be used for these purposes.

Summative Evaluation

Data Organisation

Each of the 25 research project reports from the New Basics Research Program was reviewed and summarised in terms of introduction (or purpose), data sources, methods, results and interpretation. The methods section contained comments on, as well as a description of, methods. The interpretation was constructed so as to represent the findings that could be reasonably sustained on the basis of the results taking into account the methods. These summaries were then organised around the following six themes: Rich Tasks and student work, teaching and learning, assessment, student achievement, school organisation and structure, and understandings and responses. For each of the themes a summary interpretation was developed.

Integration of Research Results

Since the New Basics Research Program was made up of 25 research reports, an approach to integrating and summarising those findings was important. In social research two approaches to research syntheses or research integration are adopted. The first is narrative integration and the second is meta-analysis or quantitative research synthesis.

Narrative methods of research integration have been the basis of traditional literature reviews. The process typically involves a series of judgements and interpretations by the reviewer and because the basis for those judgements is not explicit, differences in interpretations are hard to reconcile. The process provides opportunity for judgements about quality but encounters difficulties when the volume of literature in a field is large and variable. When the volume is large the procedure is often to include only some, judged to be the most important, studies. One approach to improving these traditional approaches has been to adopt explicit strategies for selecting and categorising studies (Light & Pillemer, 1984). In the present evaluation the number of research reports to be integrated was not large so that traditional narrative methods were practical.

Quantitative research synthesis, or meta-analysis, grew out of various approaches to applying quantitative methods of analysis to research findings. Essentially this amounts to applying systematically quantitative methods of analysis to the results of research studies in an area (Cooper & Hedges, 1994; Glass, McGaw & Smith, 1981). One of the important aspects of meta-analysis is deriving measures of effect size for each study.

There are two broad approaches to deriving estimates of effect size. One relies on converting results to a correlation coefficient and is applied where both variables are continuous. The other is based on converting results to a difference measure. This is typically a standardised mean difference where the independent variable is a dichotomous category and the dependent variable is continuous. For categorical data a difference in probabilities or proportions is commonly used to indicate effect size but Fleiss (1994) recommends the odds ratio as a preferable index of effect size. In the present evaluation the odds ratio has been used and for comparisons with other measures of effect size the logarithm of the odds ratio has been used². In the present evaluation results are typically in terms of a dichotomous classification; in the New Basics trial or not.

In the present evaluation the number of studies is limited and not all could be expressed in terms of an effect size. Consequently, it has used the principles of meta-analysis to estimate effect sizes on a common metric where possible but it has been necessary to also make use of other approaches to research integration.

² The log of an odds ratio is equivalent to the coefficient in a logistic regression equation and so it made sense to use it as a measure of effect in this exercise.

RESEARCH FRAMEWORK

The organisation of the research program conducted by the Queensland Department of Education and the Arts began with, and was organised around, a hierarchical set of research questions. The overarching or primary question was stated as “Is the New Basics Viable”?

Research Questions

In addition to the primary question there was a set of three secondary questions and a series of lower-order questions within each of these.

- Was the New Basics likely to lead to changes that are wanted?
 - Was there a causal link between the New Basics and student outcomes?
 - Were there changes in student performance in nature and depth?
 - Were there changes in classrooms?
- Was the New Basics likely to be accepted?
 - How did stakeholders perceive the New Basics?
 - How did stakeholders perceive the process of implementing the New Basics?
- Was the New Basics feasible on an extended basis?
 - What were the factors that facilitate or impede implementation?
 - What had to be done for the New Basics to be extended and what were the costs?

Major themes from the research

Using the hindsight that comes from looking back at the data that have been assembled over the period of the trial it is possible to see several themes that underpin these questions and about which there are data available to inform judgements. The themes are:

- Rich tasks and qualities in student work
- Teaching and learning practices
- Assessment procedures
- Student achievement
- School organisation and structure
- Perceptions and understandings

The list of themes did not quite correspond to the end points of the hierarchy of questions in the New Basics Research Program. However, they informed those questions and provided a way of organising the results and data from the research program. In addition there were issues that were of a second order that would be informed by these primary data but where the resolution depended on bringing discrete elements together and on other considerations. These included judgements about the factors that impeded or facilitated implementation, the costs of implementation, operation and extension, and issues of acceptance and extension.

Trial Schools and Comparison Schools

The focus of the New Basics Research Program was on the 38 schools in the Phase 1 trial³. Selection of these trial schools from more than 200 that expressed interest in being part of the trial was based on a range of factors including organisational capacity. The selection was not representative of the system of government schools in Queensland. Seventeen of these were individual schools and another 21 schools were members of three clusters (groups of schools in an area that planned to work together)⁴.

For several aspects of the research program, schools designated as “like” schools or “comparison” schools were selected from amongst schools that were not part of the trial. For “like” schools this was done using three characteristics of schools that were available from data held by Education Queensland: band level of the Principal’s position, the index of relative socioeconomic disadvantage (IRSED) for the school population and the proportion of Indigenous students enrolled at the school. For “comparison” schools this was done using other data and different methods. Although each of these approaches attempts to match trial schools with similar schools not in the trial, that match can never be complete and the possibility is present that trial and like schools differ in other characteristics. Comparisons over time between trial and comparison schools provide a basis for inferences about the trial that is better than static comparisons.

In some projects, characteristics of individual students were gathered and these data were included as part of a multi-variate analysis. This provides the best method of identifying the effects of the trial because it makes allowance for the possible influence of a wider range of confounding influences at individual as well as group level. Of course, there can always be other unmeasured factors that are associated with being in the trial or not and that can influence the outcomes. However, the random assignment of schools to the trial, and of students to trial and non-trial schools, in an experimental design was never a practical possibility.

Samples and Precision

The research program involved a number of approaches to sampling as appropriate to the issue being investigated. One sampling design that was used in several of the projects, and which is common in survey research, is to sample in two stages. In the first stage schools are sampled. In the New Basics Research Program this typically involved selecting trial schools and a sample of non-trial schools with similar characteristics. In the second stage elements within schools are sampled, these elements being students,

³ Originally there was to be only 15 schools in the Phase 1 trial, but numbers grew in response to the level of interest expressed

⁴ For Phase 2 an additional 21 schools were selected from amongst schools that had originally expressed interest in being part of the trial.

samples of work, classrooms or teachers. The procedure is sound but the two-stage process results in samples that have less precision than a simple random sample of the same size. As an illustration of this issue, statistics estimated from a sample of 1000 students from 50 schools are more precise than those from a sample of 1000 students from just 10 schools (Ross & Wilson, 1994; Ross & Rust, 1994). In a two-stage sample design standard errors, and tests of significance, need to be computed either by including estimates of the design effect or by using multilevel analysis techniques. Where this has not been done standard errors are likely to have been underestimated and differences of borderline significance may not necessarily be significant⁵. These issues have been noted in the context of particular research reports.

Data Sources

Data on which the 25 research projects draw are shown in Table 1. From Table 1 it is evident that some projects draw upon common data collections and that there is a range of different types of data used in the New Basics Research Program. Table 1 describes each of the data sources briefly. The data sources include interviews and qualitative observations in schools, classroom observations and classroom environment surveys, assessment data based on selected instruments or on system-wide assessment programs, samples and folios of student work, systems-wide data on opinions of schools, discourse on discussion lists and in forums and cost data. Table 1 indicates that a variety of data underpins the research program.

In Table 1 the projects are grouped around the issues that formed the basis for the chapters in this report: Rich Tasks and student work, teaching and learning, assessment, student achievement, perceptions, understandings and other matters. Some, but not all of the research involved comparing the 38 phase one trial schools compared with a group of schools with similar characteristics known as “like” schools. Other research involved examining changes in trial schools over time (sometimes compared with changes in like schools over the same time). In addition to these there were other pieces of research that examined the properties of an important element in the New Basics such as “richness” in relation to the Rich Tasks and pieces of research that describe and characterise the features of schools in the trial.

A substantive issue that is not evident in the framework is a direct measure of student engagement (as a converse of alienation), which was, arguably an important aspect of the rationale for New Basics. To some extent the issue is touched on in the enacted pedagogy assessment and in the survey of satisfaction but it would have been valuable to have a measure of the extent to which students did become engaged in working with the New Basics. A second issue concerns longitudinal change by individuals in terms of achievement and other areas. There are organisational difficulties in following the progress of individuals but growth measures tend to be sensitive to pedagogy and could have added to the perspectives obtained from the trial. Notwithstanding these issues, the projects were congruent with the research questions and themes and the data sources generally provided an adequate basis to support the research projects.

⁵ It should also be noted that the stage 1 selection involves stratification which improves precision at that level.

Table 1 Project Data Sources in New Basics Research Program**Rich tasks and qualities in student work**

Can richness be recognised in student work?	18 samples of selected student work samples from 14 schools were judged on 10 dimensions of richness by community members, experts and educators using paired comparisons methods
Putting Rich Tasks to the test	Student work from 26 folios of Rich Tasks from trial (21) and 26 folios of best work from non-trial (18) schools were assessed by 79 judges (43 educators and 36 others) in two full-day meetings.
School staffs' responses to implementing Rich Tasks	Interviews in 13 (of 38) trial schools with principals, deputy principals, critical friends, teacher aides and at least two teachers per school. Audio records of interviews were transcribed and analysed. The interviews did not explicitly solicit information about Rich Tasks but many respondents volunteered comments about the Rich Tasks.
Student work: What we got, what we saw	Samples of student work from 22 schools (62 students) in 2001, 24 schools (119 students) in 2002 and 57 schools (329 students) in 2003.
Show or substance: Has New Basics had real impact on "school"?	41 teachers from 13 non-government schools judged 73 folios against six domains; Analysis of texts held by the NB branch for evidence of impact.

Teaching and learning

Looking for change in teacher practice	Observation schedules from the QSRLS were used to code classroom practice in trial schools over 2001 (6), 2002 (13) and 2003 (13) and six non-trial schools in 2003. A total of 326 lessons were observed using a high-inference 20-item schedule. Four coders were used and inter-coder reliabilities were 0.70 or higher.
Student perspectives on classroom experience	A 51-item on-line survey instrument (the enacted pedagogical assessment) was used to obtain students' perceptions of teaching practices in their classrooms. The scales correspond to most of those (18 of the 20) in the QSRLS. The achieved sample consisted of 2336 students from 29 trial and 32 matched "like" non-trial schools.

Table 1 Project Data Sources in New Basics Research Program (continued)

Assessment

Teachers' conceptions of assessment	Questionnaire administered as part of a larger survey of 1525 teachers from 92 schools in 2003
Rich Task assessment: How teachers make judgments	Three sources of data were used: comments from 12 (out of 17) moderators, pattern analysis of annotated grading masters and teacher evaluations of student work against features.
Rich Task assessment system	Comparison of implemented system with model specified and other moderation models based on features of good assessment systems.
Rich Task assessment: Profiles of results	Based on records of rich task grades from 26 trial schools at Years 3 and 6 and 12 schools at Year 9. Analysis of rich task grades from (non-special schools), dimensionality of RT achievement and participation indexes.

Student achievement

International Schools' Assessment	In February and November 2003 Phase 1 trial schools (24 primary and 10 secondary) and a selection of "like" schools (10 at Year 3, 13 at Year 6 and 11 at Year 9) sat for the International Schools Assessment in reading, mathematics and science. In March there were 1443, 1740 and 4444 students at each year level. In addition a number of non-government schools (18, 22 and 14) and their students participated.
Aspects of Literacy & Numeracy	Based on system-wide data on literacy and numeracy assessments at Years 3, 5 and 7 with a focus on comparison between trial and non-trial schools
World Class Tests	Based on the WCT problem solving tests April and July of 2002 and March 2003. In 2002 data were gathered from 37 schools in April and 46 schools in July (out of 59 phase 1 and 2 trial schools) involving 195 and 216 9-year olds and 152 and 302 13-year-olds. In March 2003 38 trial and 24 "like" non-trial schools, with 238 9-year-olds and 521 13-year-olds participated in a comparative study. In addition 321 students from 25 non-state schools participated in the March 2003 WCT.

Table 1 Project Data Sources in New Basics Research Program (continued)

School organisation and structure	
Teacher professional communities and New Basics	Questionnaire surveys of 868 teachers in 2002 and 1525 teachers from 92 schools in 2003
School structural change and New Basics	Interview study in 13 trial schools in 2003
School staff's views of sharing ownership of reform	Interview study in 13 trial schools in 2003
School staff's view of the external community's role	Interview study in 13 trial schools in 2003
Role of critical friend	Discussion day narrative and questionnaire to 30 critical friends, 33 principals, 27 teachers and 23 central office staff
Perceptions and understandings	
Changes in school satisfaction measure	Secondary analysis of EQ school opinion survey in each year from 2000 to 2003
The construction of knowledge about New Basics	Interviews with school staff regarding where they obtained information about New Basics together with analysis of documents and observations in eight trial schools (phase 1 and phase 2). An analysis of the congruence between messages and site-constructed knowledge was conducted.
The New Basics discussion lists: What was said, what was heard	A descriptive analysis of the interactions between participants in three electronic discussion lists over different periods of time.
Stakeholder responses to the New Basics	Based on a synthesis of records of community consultation, forums, and market research based on telephone surveys, personal interviews and focus groups.
Other matters	
What did the New Basics Trial cost and was it value for money?	Application of cost analysis protocols to official data about costs of the NB trial.
What will it cost for the New Basics to be extended?	Desk analysis of cost of extension using existing cost frameworks.

There were also unintended outcomes to be considered such as the potential role of the Rich Tasks and the assessment system including moderation in promoting change in other aspects of schooling. The sequence of changes in each of the elements as well as the magnitude of the changes in what happened suggest that the Rich Tasks and the moderation system have been important to the implementation of the New Basics. The role of appropriate assessment systems in promoting changes in teaching and the enacted curriculum has been noted in other approaches to system-wide change (Earl, 2003; Earl, 2004).

Summary

The New Basics Research Program was structured around key areas that were central to the New Basics. There were projects concerned with the analysis of data about achievement as well as other outcomes. There were also projects that addressed issues concerned with teaching and learning in schools and classrooms. A number of projects address issues concerned with the adoption and implementation of new basics in schools as well as the role of the systems in that implementation. The larger issues concerned with the implications of implementing New Basics more widely need to be addressed by integrating the results from the individual research projects across themes rather than within themes. Those “large change” questions are important for a wider understanding of systemic change as well as for the particulars of this program.

RICH TASKS AND STUDENT WORK

Rich Tasks were a form of authentic assessment but they also had a purpose in shaping teaching and learning. From the outset they were seen as central to the New Basics. The *New Basics Project Technical Paper* (Education Queensland, 2000) indicated that Rich Tasks were designed to be both “curriculum planning and student assessment”, to “provide balanced and comprehensive coverage of the four categories of New Basics”, be “representative of an educational outcome of demonstrable and substantial intellectual and educational value” and be “problem based, with relevance to and power in new worlds of work and everyday life”. They were also to be “sufficient in developmental, cognitive and intellectual depth and breadth to guide curriculum planning across a significant span of schooling”. Each Rich Task drew on specified skills (called repertoires of practice and defined as the cognitive and cultural, linguistic and social skills needed to complete the task) and fields of knowledge (called operational fields needed to complete the task)

Table 2 Summary of Rich Tasks for Years 1 to 3

Years 1-3	
1 - Webpage Design	Students will collect information about themselves, their school and their community. They will use this information to design web pages in their websites and respond to questions electronically.
2 – Multimedia Presentation of an Endangered Plant or Animal	Students will investigate a threatened Australian plant or animal and the extent to which it is at risk. They will use this investigation to take constructive action and create a persuasive and informative multimedia presentation.
3 - Physical Fitness	Students will memorise, rehearse and master dances of different forms. They will prepare introductions for their performed dances by investigating the role of dance and the cultural context of their dances. They will measure and monitor their fitness as they engage in a high level of physical activity.
4 - Read and Talk About Stories	Students will view, read and listen to fiction stories presented in different media forms. They will analyse characters and settings and compare different stories and different media, incorporating their own experiences. They will present their ideas in a performance using a selected combination of words, visual images, music and drama.
5 - Historical and Social Aspects of a Craft	Students will explore craft as a personal, social and cultural endeavour. They will prepare and run a stall that showcases a chosen craft and an object or objet d'art that they have made as an example of that craft.

Source: Adapted from New Basics: The why, what, how and when of rich tasks (Education Queensland, 2001c).

Table 3 Summary of Rich Tasks for Years 4 to 6

Years 4-6	
1 - Travel Itineraries	Students will design alternative itineraries of interest to a party comprising the student and an exchange student, and to be accompanied by an adult. They will identify a range of issues including transport options, tourist attractions and sites of historical and cultural significance. They will present costings and reasons for their choices.
2 - Narrative Text: Away with Words	Students will critically examine books written for emergent readers. They will determine the criteria for categorising these books and select one category for further examination. In this selected category, students will present a review of a book. Using the selected category, they will then choose an aspect of nature and create an illustrated storybook - crafted by hand and/or electronic technology - for their peers to review.
3 - Personal Health Plan	Students will identify and understand an aspect of their personal health and fitness and, on the basis of this, will develop and implement a plan for improving this aspect. By collecting, organising and presenting data, they will evaluate the extent to which the goals have been achieved and the contribution of factors to this improvement.
4 - A Celebratory, Festive or Artistic Event or Performance	Students will work within teams, in different capacities, in planning, organising, creating and performing in a celebratory, festive or artistic event or performance that takes place at or outside the school.
5 - Oral Histories and Diverse and Changing Lifestyles	Students will explore change in, and diversity of, twentieth century lifestyles, with particular reference to the nature of work, by recording oral histories from various members of their own community, including people in a variety of cultural groups. They will use the oral histories as the basis for a finished electronic media presentation that portrays significant changes in work practices in the past and predicts how work practices might change in the foreseeable future.
6 - Design, Make and Display a Product	Students will design, or improve the design of, a purposeful product. They will make the product or a working model or prototype. As part of a public display promoting their product, they will flesh out a (restricted) marketing plan and explore the suitability of materials for mass manufacture.
7 - Space Futures	Students will engage with the exploration of space and with the techniques and procedures of the mathematical and physical sciences. They will produce a model of the solar system, investigate the impact of space travel on life on Earth, and produce a coherent design for an experiment to be performed on a spacecraft.

Source: Adapted from New Basics: The why, what, how and when of rich tasks (Education Queensland, 2001c).

There were three reporting points: at the end of Years 3, 6 and 9. The set of 20 Rich Tasks covering the three three-year spans is summarised in Tables 2, 3 and 4. Each of the sets of Rich Tasks was intended to cover a balance of fields of knowledge and generic skills. The tasks were published as a teacher booklet and distributed to trial schools to form the basis of how schools reported to parents and the system on student achievement. Rich Task assessment, and associated moderation, was envisaged as a key element of *New Basics* and as a major driver of changes in other elements of *New Basics* (Education Queensland, 2001a: 4).

Table 4 Summary of Rich Tasks for Years 7 to 9

Years 7-9	
1 - Science and Ethics Confer	Students will identify, explore and make judgments on a biotechnological process to which there are ethical dimensions. They will identify scientific techniques used, along with significant recent contributions to the field. They will also research frameworks of ethical principles for coming to terms with an identified ethical issue or question. Using this information, they will prepare pre-conference materials for an international conference that will feature selected speakers who are leading lights in their respective fields.
2 - Improving Wellbeing in the Community	Students will work with a local community to develop a plan for improving an aspect of the wellbeing of this community and then enact the plan, modifying it as necessary. They will evaluate the level of success they experience in enacting their plan and, where necessary, recommend future actions.
3 - The Built Environment: Designing a Structure	Students will identify a client's needs and take these and other factors into account in preparing a design brief for a structure. They will design an environmentally sensitive and aesthetic structure to fulfil this brief and communicate the design through sketches, plans and models. They will give due consideration to structure and materials, quantities and costs.
4 - Australian National Identity: Influences and Perspectives	Through the creation, production and presentation of a powerful, filmed documentary that incorporates information gleaned from research and interviews with people from different cultural backgrounds, students will demonstrate knowledge and understanding of different influences and perspectives on "Australian national identity".
5 - Personal Career Development Plan	Students will undertake a career planning process. They will describe features of a range of work options and their associated expectations; assess their own existing strengths, interests, achievements and areas to be developed; identify potential careers; and produce an individual career development plan, including an up-to-date résumé.
6 - Opinion-making Oracy	Students will make forceful speeches on an issue of international or national significance to three unlike audiences in different forums.
7 - Pi in the Sky	Students will demonstrate an understanding of different mathematical approaches used to frame and answer questions about astronomy asked by cultures from three different historical ages. For each culture, they will immerse themselves in one such question as well as the ways in which the culture used or developed mathematics to frame and answer the question. They will then present one of three lessons, chosen at random, to communicate the essential ideas and techniques of the mathematics of the situation.
7 (alternative) – The Shape We're In	Students use high mathematical skills in number, measurement and spatial concepts to investigate alternative shapes and/or dimensions for at least one container, one domestic object, one mechanical device and the structure of an object from nature. They will then adapt one of these objects, identify the purposes and consequences of the adaptation and explore the mathematics involved.
8 - International Trade	Based on knowledge of the way in which international trade occurs and is reported as well as knowledge of the needs and wants of another culture, students will identify and provide a detailed analysis of an export opportunity. They will take advantage of their skills in a language other than English to present a talk and supporting literature to promote this export opportunity to different buyers and backers.

Source: Adapted from New Basics: The why, what, how and when of rich tasks (Education Queensland, 2001c).

The Concept of Richness

Rich tasks were argued to be a “rigorous intellectual focus for student work” that simplified the crowding and diffuse character of the school curriculum (Education Queensland, 2001c: 5). They were seen as a culminating performance that required the development of important intellectual skills and was transdisciplinary (drawing on a range of disciplines but retaining the integrity of each) (Education Queensland, 2001c: 5).

Subsequently, as a synthesis of descriptions in the earlier documents and the results of consultations, richness came to be defined in terms of three broad dimensions (Matters, 2004):

- Intellectual engagement manifested in deep thinking and understanding, going beyond (interpreting) the data that are presented, and giving consideration to aesthetics in presentation.
- Engagement in discipline-based, and trans-disciplinary, learning by demonstrating deep and coherent knowledge of a field, drawing on knowledge from several fields and drawing on skills and practices across fields.
- Engagement in problem solving and action by identifying and analysing problems, making decisions based on information and discussion and engaging in social action and communicating in different forms.

Each of these three dimensions was elaborated in terms of ten elements that had been identified through consultations with judges working with samples of student work. The pattern has been illustrated in Figure 1.

The elements making up the three broad dimensions were as follows:

- cognitive depth, attention to aesthetics, interpretation beyond presented data,
- knowledge from several fields, incorporating practices across disciplines, coherence of knowledge, and
- significance, collaboration, judgement, action orientation.

The dimensions and the constituent elements were properties of the student work, although they were sometimes expressed as what was inferred about the process that resulted in the sample of work. In a few instances the elements did not sit readily with the dimensions. For example attention to aesthetic consideration did not unambiguously fit with intellectual engagement even though it was reported to be often a component of the judgement of richness in student work.

Construct Validity: Recognition of Richness

It was claimed that the Rich Tasks promoted higher quality work from students. An important question therefore was whether different people could identify ‘richness’ in similar ways. One of the studies in the New Basics Research Program investigated whether a range of people, from within and outside education, could consistently identify richness in student work.

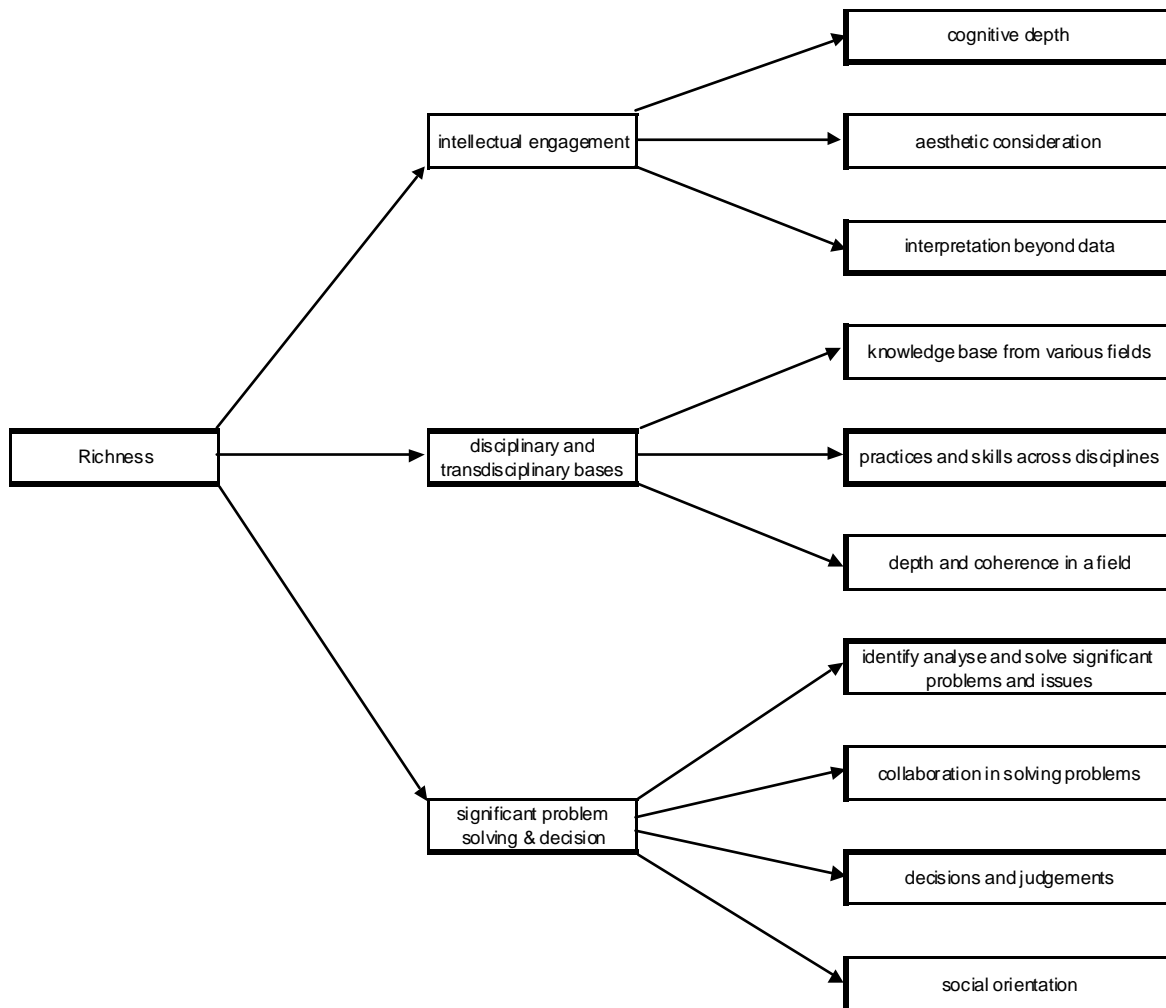


Figure 1 Dimensions and Elements of Richness in New Basics

Sources

Eighteen samples of student work were chosen from an exercise in which teachers were invited to provide a box of student work (including some that was related to the Rich Tasks) from Years 6, 7, 8 and 9. These work samples were from 14 (trial and non-trial) schools and were chosen to represent a range of degrees of richness. There were difficulties in obtaining samples of student work possibly because of reluctance by teachers to provide these samples. The difficulties in sourcing student work precluded a rigorous sampling procedure and this resulted in less precision than desirable and possible sample bias.

Method

A total of 57 judges (including 30 educators and 27 other community members) were involved in assessing these samples of student work. Initially the judges were provided with little guidance about the concept of richness but asked to make pair-wise comparisons of four pairs of student work samples. They then participated in focus group discussions and provided input on the meaning of the term “richness” in the context of

New Basics. This was followed by a training session. Finally they each made pair-wise comparisons of four new sets of work samples.

Results

From the written comments, and in the discussion, it was reported that community members viewed “richness” in more diverse and externally oriented ways than did educators. In general it was found that the judges were making reasonably consistent judgments about most folios, but quite disparate judgments about a small number of folios. Training made a small difference to the consistency of judgements. Different groups of judges expressed their understanding of richness in ways that were broadly congruent with the New Basics construct of richness but community members’ seemed to capture more of the real-world connectedness of richness while the educators tended to express themselves in more education-based terms.

Interpretation

The analysis suggested that it was possible to make valid judgments about the richness in student work. This was congruent with a view that a range of people including those without specialist educational expertise could identify “richness” (as defined in New Basics). Although the reliability of judgments was improved slightly after training, it was evident before training. There was broad agreement between their articulation of richness by the judges and the construct of richness expressed in New Basics documents.

Comparisons of Rich Tasks and Other Samples of Student Work

Source

All trial schools (other than special schools) were invited to nominate two good examples from designated Rich Tasks at each juncture level (Years 3, 6 and 9) that had been completed in term 3 of 2003. Thirty-one of the 35 invited schools submitted work. Within the work submitted there was a further selection based on completeness and range so that 26 of the 72 examples provided were used in the research activity. It is not clear whether these criteria for selection also removed less rich samples of student work. Non-trial schools were selected on the basis of being schools where students had records of high achievement in other areas. Each nominated two completed pieces of assessment at each of Years 3, 6 and 9 in term 3 2003 from high achieving students within the school. Twenty-six folios from non-trial schools were selected for the research activity. It is relevant that the comparison samples from non-trial schools were with work from high achieving students in schools that had records of high achievement. However, it also seems that that the work from trial schools may have represented better work from those schools. Thus the selection procedure resulted in a potential threat (of uncertain magnitude) to the validity of the comparisons.

Method

A group of 79 judges was selected from community (non-educators) members (36) and educators (43, 22 of whom were from trial schools⁶) to assess the tasks in terms of richness. The judges followed three steps. First, with little guidance the judges they

⁶ The judgements by judges from New Basics trial schools were similar to those from other schools.

assessed which of two samples in a pair was richer⁷. Secondly, they participated in focus groups and articulated their conceptions of richness. Thirdly, they conducted other sets of paired comparisons in relation to ten specific elements of richness, summary statements of three broad dimensions and overall richness.

Results

The analysis of the final judgements indicated that at Year 6 the Rich Tasks were judged to be richer than other student work at that Year level. Data are recorded in Table 5 which shows mean ranks and the probability that a folio from a trial school would be ranked higher than a folio from a non-trial school (described in the table as the rich task effect). If the rich task effect was 0.5 the probability was equal, if it was greater than 0.5 the trial school work was more likely to be richer and if it was less than 0.5 the trial school work was likely to be less rich. In meta-analysis the difference in proportions is sometimes taken as an index of effect size (Cooper and Hedges, 1994: 234-239). In this case the difference between actual probability and equal probability at Year 6 would be 0.24. At Year 3 and Year 9 there was no significant difference between the Rich Tasks from trial schools and student work from non-trial schools (with differences in probabilities of -0.16 and 0.01 respectively). Fleiss (1994) recommends using the odds-ratio as a measure of effect size for cross-classified categorical data (even though it is not strictly comparable to a mean standardized difference)⁸. Odds-ratios for trial compared to non-trial schools were 0.51 at Year 3, 2.85 at Year 6 and 1.04 at Year 9. The Rich Task effect value of 0.51 at Year 9 indicated that Rich Tasks at that Year level were judged to be of similar richness to work from high achieving students at non-trial schools. The value of less than 0.5 at Year 3 indicated there was a tendency for student work by high achieving students in other schools to be richer than the Rich Tasks from trial schools, even though the difference was not statistically significant.

Table 5 Rankings of Richness in Student Work

Year	Rank mean		“Rich Task effect”	Odds Ratio Trial: non-trial	W_N^{BF}
	Trial	Non-trial			
3	7.3	9.7	.34	0.51	-0.848
6	10.9	6.9	.74	2.85	1.800
9	10.1	9.9	.51	1.04	0.079

Source: *New Basics Research Report Number 2*

⁷ The technique of pair wise comparisons is a well-established procedure for assessment concerned with complex concepts (see David, 1988).

⁸ Odds ratios are the ratio of the odds of a member of one group attaining a given criterion rather than not attaining that criterion to the odds of a member of another group attaining the same criterion rather than not. Odds ratios are always positive. An odds ratio equal to one signifies no effect of the variable concerned on the criterion. Odds ratios above one indicate an increased likelihood of attaining the criterion and odds ratios below one indicate a decreased likelihood. The further an odds ratio is from one, the stronger the effect of the variable. This measure has several desirable properties. First, it uses more information in describing the association between two variables and thus is a more complete measure. Second, it has the property of ‘marginal invariance’, that is, the strength of an association between two measures is not affected by their distributions. Third, odds ratios can be used to interpret the effects of factors on participation in a multivariate context.

Interpretation

One of the questions that arose from these results was the extent to which the judgements of richness were a consequence of the principle that to “be truly rich, a task must be transdisciplinary”. The nature of most Rich Tasks required perspectives from a range of fields whereas other samples of student work could have been from one field. In other words was the work in the Rich Tasks different from other student work in elements other than those reflecting a transdisciplinary approach? In some senses the question did not matter because richness had been defined as incorporating the drawing together of knowledge and perspectives from several fields. If this were accepted then the term “richness” would not be taken as synonymous with intellectual quality, which could arguably be evident in work from just one field. Of course there was support from the judgements made before training that “richness” was not simply a reflection of transdisciplinary elements. Furthermore there was evidence that all the elements of richness (including transdisciplinary qualities) were strongly intercorrelated.

The different pattern of results for Year 6 compared to Year 3 and Year 9 invited further interpretative comment. Such a pattern could have reflected differences in the nature of the Rich Tasks provided at each of the juncture years, differences in what happened in other school activities at those years or differences in context that influenced implementation of Rich Tasks. One interpretation could be that Year 6 was the most fertile context for the adoption of Rich Tasks and that this was reflected in the nature of the work done by students. Students might have developed the proficiencies needed to tackle Rich Tasks by Year 6 and organisational structures in Year 6 might have been more conducive to these forms of work. At Year 3 the focus of teaching could have been on the development of foundation skills for which traditional forms of teaching were seen as appropriate and without which working on the more broadly conceived Rich Tasks was difficult. For the Year 9 juncture, the three preceding years spanned the last year of primary school and the first two years of secondary school. This could have impeded the full development of the Rich Tasks. In addition the emergence of more specialised patterns of interest, as well more segmented teaching traditions and organisational structures in secondary school could have acted as an impediment to a development such as Rich Tasks. Twelve of the 38 trial schools enrolled secondary students so that the sample of schools was small and generalisation limited. Given these constraints, and the condition that the comparison was with work from high achieving students in other schools, the Year 9 results suggest that extended work tasks that could be accommodated better in high schools might also have effects on richness.

Teachers’ Responses to Implementing Rich Tasks

Two data sources were used for this study. One was the 13-school study, from which teacher interviews provided data for an analysis of teachers’ perspectives on planning, implementing and assessing Rich Tasks⁹. This involved considering the role of Rich

⁹ The 13-school study is described later in the report. Thirteen of the trial schools were selected as case studies taking into account characteristics that were likely to be related to variations in factors that would facilitate the development of school professional learning communities: school size, location, school type, socio-economic demographics, cultural factors, participation in clusters, critical friend usage, and administrative approach. Six of the schools were in metropolitan locations, four were in regional locations, two were in rural locations, and one was in a remote area. There were three high schools, eight primary schools (one with P to 10 enrolments), a special school and a school of distance education. In each school principals, deputy principals, critical friends,

Tasks in backward mapping for three-year spans, accountability, connectedness to the wider world and connections across disciplines. A second data set was the records and reports from the Rich Task review in August and September 2003.

Instrument

One instrument was the open interview schedule used in the larger 13-school study with the procedure for analysis being the same as reported for that study. The other source of data (not really an instrument) was the records of the Rich Task consultations.

Method

It is important to note that questions about Rich Tasks were not part of the interviews even though teachers made comments about Rich Tasks during the interview. It is those unsolicited comments that provided the basis for the analysis that followed. The records of the Rich Task review were not a systematic data collection but provided an opportunity to use material intended for another purpose to inform these deliberations. This is an important caveat because the absence of comment does not mean that an issue was not of concern.

Results

The nature of the data that were used in this investigation meant that there was not an explicit distinction between results and interpretation. This report of results depended upon the interpretations made by researchers during the process of analysis. Many teachers reported involvement in planning Rich Tasks as a curriculum for three years but that planning for three years was not fully appreciated at first. It appeared that teachers became more aware of three-year planning requirements as a result of moderation. There was also evidence that teacher interpretations of requirements differed from moderator expectations. The transdisciplinary character of Rich Tasks was recognised but implementation was incomplete especially in high schools. Although it was recognised that Rich Tasks should link students' work with the world outside the school, this had not been attained in the ways that had been envisaged or on a widespread basis across trial schools.

Interpretation

The evidence in this research report suggested that although the purposes of the Rich Tasks were reasonably well understood they had not been implemented with complete fidelity with respect to using them as the basis for three-year planning. Studies of school reform recognise that local interpretation and adaptation are part of implementing wider initiatives (Cuban, 1998). In this case there was evidence that the moderation process focused teacher attention on, and made more explicit, the requirements of New Basics. It seemed that teachers became actively engaged in the tasks even though they experienced difficulty in doing them.

teacher aides, and a minimum of two teachers were interviewed. These interviews were held early in the third year of the trial. All the interviews were transcribed so that they could be analysed systematically.

Samples of Student Work

Experience of New Basics was intended to influence the intellectual quality of students' everyday work¹⁰. There was an analysis of student work collected from trial and non-trial schools. It is important to note that this research project examined student work within each 3-year span rather than in the culminating Rich Tasks.

Sources

Student work was sampled from Years 7 and 9 in 2001, Years 2, 4, 7 and 8 in 2002 and Years 2, 5, 8 and 9 in 2003. In 2001 schools were requested to provide work that three students had produced during a nominated three-week period. Although the schools made the selection of the students the request was for "the maximum amount of evidence of what these students had produced". This resulted in work from 62 students in 22 schools. In 2002, the request was explicit in referring to "three of the best students who were likely to produce work of high intellectual quality in the classes of the best teachers". In this year there was work provided from 119 students in 24 schools and in 2003 there were samples of work from 329 students in 57 schools. In each year the work samples came from trial and non-trial schools. At least in 2002 and 2003 the sampling of student work highlighted exemplary rather than typical work and this limits the conclusions that can be drawn about typical work in classrooms.

Method

Student work was coded according to criteria concerned with range of media, depth of content, coherence, connectedness and intellectual quality. Coders were trained in the meaning of terms on the coding sheets. Intellectual quality was used as a global overall construct encompassing the specific constructs. In 2001 one coder assessed each piece of work (with recoding being conducted in 2002) and in 2002 also one coder (from a panel of nine) assessed each piece of work. In 2003 at least two coders (from a panel of 23) coded each piece of work with any differences being resolved in favour of the higher rating. In this year there was between 56 per cent and 87 per cent agreement between coders. In addition to the coding, detailed documentation was recorded of the characteristics of work that was deemed to be of high or low intellectual quality.

Results

The work collected in 2001 provided little evidence of intellectual quality. It emphasised basic knowledge and skills across subjects and there were few connections to the real world. There were few statistically significant differences between the work from trial and non-trial schools. Similar results were found in 2002 and there was no evidence of improvement in the work being done in trial schools in 2002¹¹. In 2003 there was a greater number of significant differences in connectedness and coherence with folios from trial schools revealing higher scores than in previous years and higher scores than

¹⁰ Student work has not been widely used in research. It has been used in the *Schools Around the World* project to investigate differences in expectations between countries (Stoel & Anderson, 2000) and in the evaluation of comprehensive school reform projects (Bryk, Nagoada, & Newmann, 2000; Berends, Bodilly, & Kirby, 2002).

¹¹ Work from trial schools contained more performing arts, physical activity and group work than that from non-trial schools and work from trial schools in Year 8 had a stronger focus and wider content than work from non-trial schools.

for non-trial schools¹². In terms of intellectual quality the percentages at Year 9 were 84 per cent of trial schools and 58 per cent of non-trial schools (an odds ratio of 3.8). At Year 8 the corresponding figures were 69 per cent and 45 per cent (an odds ratio of 2.7)¹³. The preceding differences were statistically significant. At Year 5 and Year 2 the differences were just one percentage point (odds ratios of 0.96 and 1.04 respectively).

Interpretation

There was little evidence of differences between trial and non-trial schools in the intellectual quality of student work during 2001 and 2002. The results from the analysis of student work in 2003 provided some evidence of a change in the overall intellectual quality of student work from trial schools in Years 8 and 9. The changes that were observed at those levels, and at Year 4, arose from greater connectedness (links to the wider world). Differences were not evident in either greater depth of content or range of content being present in work from trial compared to non-trial schools. Two of the possible explanations for this change in 2003 were that changes of this type took time to be fully implemented and that change was more likely to occur when there was a moderated assessment system in operation. The moderated assessment system was visible within the educational community and made expectations of standards explicit.

Impact on School Work

New Basics was intended to change student learning. This research project investigated whether the work students did manifested the intentions of New Basics.

Sources

There were two distinct sources of data. The first consisted of folios of student work from the “Richness Project”, the “Student Work Project” and the Rich Task ratification part of the moderation process. The second consisted of files held in the Assessment and New Basics branch from school files (including school curriculum plans), submissions from stakeholders to the Rich Task Review, Productive Pedagogies research notes, the construction of knowledge project, school visit and rich task reports of moderators and notes from the check conducted in late 2002 in all trial schools to ascertain the nature and extent of implementation of Rich Tasks.

Method

In the analysis of the first data source, 41 teachers from 13 non-state schools participated in a two-day process where panels of teachers examined 73 folios of student work against criteria relating to scope, coverage of skills for the future, coverage of transferable skills, the presence of New Basics referents, attention to more than superficial features and evidence of deep understanding. The teachers worked in 11 panels of three and two panels of four. For each folio the panels voted on whether there was evidence of the

¹² There were some statistically significant differences between connectedness ratings for trial and non-trial schools in Years 8 and 9. In terms of connectedness in Year 9 the percentages were 84% for trial schools and 25% for non-trial schools, at Year 8 the percentages were 58% and 3% and at Year 5 the figures were 62% and 23%.

¹³ Odds ratios are the ratio of the odds of a member of one group attaining a given criterion rather than not attaining that criterion to the odds of a member of another group attaining the same criterion rather than not. The index is explained in greater detail in a previous section of the chapter.

specified New Basics principles. Work was considered to manifest New Basics principles if it was considered to contain four of the six principles. Forty-six of the 73 folios were viewed twice. The analysis of the new Basics texts was less systematic and involved a trawling for evidence by staff in the New Basics branch.

Results

Although the incidence of New Basics characteristics in the box of student work folios was higher for trial schools compared to non-trial schools it was only present in 21 per cent of cases. Across the three sources of work, manifestation of New Basics was highest in work from the Richness project, although only in 41 per cent of cases. New Basics characteristics were low in work from the Moderation process even though these work samples formed part of a Rich task; only 32 per cent.. The evidence in the New Basics texts was less compelling than that found in the student work.

Interpretation

Overall the analysis of student work supported the conclusion that in the New Basics trial, there were changes in practices that had been intended. Rich Tasks had wider and deeper scope than regular projects, connectedness and transferable skills were given attention, student work was valued for more than superficial features. The analysis of the texts is less clear.

Summary

The concept of richness that underpinned the Rich Tasks incorporated intellectual engagement (in thinking, understanding, interpreting data), engagement with knowledge from more than one field (used in a way that was deep and coherent) and engagement in problem solving (with an action orientation). Although the concept was multi-faceted, it was found that both educators and others could recognise richness and largely agree about its presence. There was consistency between their views of richness and the description of richness in New Basics.

When Rich Tasks from Year 6 students in trial schools were compared with work from high achieving Year 6 students in other schools, it was found that there was a higher level of richness in the work from trial schools. Notwithstanding that part of this could have arisen from the transdisciplinary requirement, it represented a significant accomplishment that was evident in each of the elements of richness. There was a tendency that was not statistically significant, in the Year 9 work for the Rich Tasks from trial schools to be richer than work from high achieving students in other schools. Therefore one cannot be sure that that such a difference did not arise by chance. In Year 3 there was no difference at all between trial and non-trial schools. Analyses of regular student work in years other than the juncture years, or in earlier stages of the trial, showed more limited and inconsistent evidence of effects on overall intellectual quality. However, there was evidence of stronger “connectedness”¹⁴ in work from trial schools in Years 8 and 9.

In general the patterns suggested that an assessment system impacts on the work done by students by making expectations explicit and by providing recognition for the implementation of tasks in the manner intended. Although the purposes of the Rich

¹⁴ Connectedness is one of the elements of intellectual quality.

Tasks were understood in principle they had not been implemented with complete fidelity. The moderation process appeared to clarify the requirements for the Rich Tasks.

TEACHING AND LEARNING

New Basics aimed to enhance the pedagogical approaches used by teachers. The enhancement of pedagogy was pursued through the implementation of the New Basics Framework and by direct in-service training courses for teachers in Productive Pedagogies. Two research studies examined changes in teaching and in learning. The first was based on observation of teaching practices in classrooms and involved both a study of changes in a set of trial schools over time and a comparison of trial and non-trial schools. The second was based on the perceptions of students on what happened in their classrooms in samples of trial and non-trial schools.

The background to the teaching practices that were intended to be part of New Basics came from a study conducted in the United States by the Centre on the Organization of Restructuring of Schools (CORS) (Newmann & Wehlage, 1993; Newmann, Marks, & Gamoran, 1995). It had concluded that a focus on student learning, authentic pedagogy, school organisational capacity and external support could result in improved learning. Authentic pedagogy was based on the extent students were challenged to think, develop understanding, and apply academic learning to real and important world problems (Newmann et al., 1995).

Instruments from this study were adapted and used in the Queensland School Reform Longitudinal Study (QSRLS) (University of Queensland, 2001). In the QSRLS the term “Productive Pedagogies” was used to describe a set of classroom strategies that teachers could use to focus instruction and improve student outcomes. These strategies were based on the authentic pedagogy model but extended to encompass the extent to which classroom environments were supportive and the recognition of difference. These strategies are themselves conceptually complex and multi-faceted. Hence, they are described here as elements rather than strategies. There were four dimensions underpinning the 20 elements: intellectual quality; connectedness; supportive classroom environment; and recognition of difference. The intended connections between the elements and the dimensions were as shown in Table 6.

Table 6 Dimensions and elements of productive pedagogies

Dimension	Elements
Intellectual quality	Higher-order thinking, metalanguage, substantive conversation, knowledge as problematic, deep knowledge, deep understanding
Connectedness	Knowledge integration, (links to) background knowledge, connectedness to the world, problem-based curriculum
Supportive environment	Student control, social support, engagement, explicit criteria, self-regulation
Recognition of difference	Cultural knowledge, inclusivity, narrative, group identity, citizenship

The Classroom Observation Study

In the classroom observation study a selection of classrooms was observed using the same observation schedule as had been used in the QSRLS. The study was conducted in a small number of schools (see discussion of method) with teachers selected by those schools who were believed to use a range of approaches. Consequently the selection was not systematic and this limits the generalisations that can be inferred. Each observation point corresponded to the one of the 20 elements of the productive pedagogies framework. Trained observers recorded what they saw and perceived on a five-point scale for each element according to frequency or level¹⁵.

Instrument

In the terminology used to categorise classroom observation tools this would be called a high inference tool because observers are required to make inferences or judgements on the basis of what they see rather than simply record an unambiguous observation. The successful use of high-inference tools is contingent upon the training of observers to an even greater extent than low-inference tools. Three trained observers made the observations. These three observers coded 25 lessons in 2003 (23 by two observers and two by three observers) so as to assess the agreement between observers¹⁶. The level of agreement was estimated using the weighted value of a non-parametric correlation coefficient (Agresti, 2002). Although the extent of agreement varied across elements the median value of the coefficient was 0.72, which would be considered to be a moderately sound level of agreement (although it reflects some variation between observers)¹⁷. This suggests that the observers were assessing lessons using similar underlying dimensions. An index of the consistency of ratings by each observer across the elements making up each dimension was also calculated (coefficient alpha). It ranged from a highly satisfactory level of 0.88 for intellectual quality to an adequate 0.69 for supportive environment and recognition of difference.

A total score for each dimension was calculated as the mean of the ratings for each element making up that dimension^{18,19}. Those scores were then used for the overall analyses. There was in the pattern of correlations among the four scales (those coefficients ranged from 0.45 to 0.65) a suggestion that there may be a single second-order dimension underpinning the four dimensions used in the analyses.

¹⁵ Frequency descriptors are in increasing order from none to all. Level descriptors are from low to high (1 – 5). For the element higher-order thinking, a score of 1 corresponds to “Students are engaged only in lower-order thinking”, while a score of 5 corresponds to “Almost all students, almost all of the time, are engaged in higher-order thinking”.

¹⁶ Technically this is known as the inter-rater reliability.

¹⁷ The values of these coefficients can be thought of as similar to reliability coefficients in which case a value of 0.7 has conventionally been regarded as satisfactory for research purposes although for individual assessment higher reliabilities would be necessary (Nunally, 1978).

¹⁸ The report does not indicate a checking of the structure of the instrument as used in this study so the assigning of elements to scales has been based on the theoretical assignment and the assumption that each element is uniquely associated with one and only one dimension.

¹⁹ There is an argument in favour of weighting the elements on the basis of coefficients derived from a confirmatory factor analysis. In my judgement this would have made little difference to the scores and the results of the simple analyses conducted here. The issue might need to be revisited in relation to more complex analyses.

Method

A total of 256 lessons from trial schools formed the basis of the classroom observation study including 120 from 2003, 84 from 2002 and 52 from 2001²⁰. Taking into account the number of lessons per teacher this amounts to approximately 30 teachers in 2003, 28 teachers in 2002 and 27 teachers in 2001²¹. There were six trial primary schools over the three years of the trial plus an additional seven trial secondary schools in 2002 and 2003²². The selection of teachers from within these schools was not random but by each school on the basis of “teachers who were likely to display an array of pedagogical practices”²³. It was reported by the researchers to be difficult to persuade teachers to participate in a classroom observation study that was seen as intrusive. In addition for logistical reasons related to the availability of coders it was not possible to schedule the observations at the same time each year or within a narrow range in any year. Each of these sampling issues represents a threat to the validity of the conclusions drawn from the results and should be taken into account when interpreting those conclusions.

The research report focuses on changes over time in the trial schools both with data aggregated across all Year levels and separately for Year 5 and Year 8. In the analysis of Year 5 and Year 8 a comparison is made with data from the QSRLS in 2000 (using Year 6 and Year 8). In my view it is not possible to make direct comparisons with the QSRLS because neither the QSRLS nor the schools in the New Basics observation study were representative samples of Queensland schools and any difference might be attributable to differences in the samples²⁴. Only about half of the trial schools were also involved in the QSRLS.

In 2003 an additional six non-trial schools were involved so that 69 lessons (amounting to an estimated 18 teachers) were observed in non-trial schools in 2003. There was an analysis of cross-sectional data for trial and non-trial schools at primary and secondary level but this was not reported because of concerns over data quality. The absence of a comparison with non-trial schools (and preferably a longitudinal analysis of non-trial schools) means that the possibility that observed changes took place in all schools cannot be dismissed.

There is an issue of unit of analysis involved in these analyses because the structure of the data is complex. For example in the longitudinal study of trial schools there were 53, 84 and 120 lessons observed in successive years but these represented approximately 27, 28 and 30 teachers²⁵. One possibility for analysis would have been to use the teacher in any year as the unit of analysis. This assumes that the multiple observations of any

²⁰ The numbers of observations for 2003 were made up of 83 in primary and 37 in secondary classrooms. For 2002 they were made up of 48 in primary and 36 in secondary classrooms.

²¹ The numbers of teachers involved are estimated on the basis of information that two lessons were observed for each teacher in 2001, three lessons per teacher in 2002 and four lessons per teacher in 2003. It would be preferable to observe more lessons for each teacher but within constraints of budget and logistics this could have compromised the number of units for analysis.

²² One of these was apparently a combined primary-secondary school.

²³ In general this may have biased the sample (in both trial and non-trial schools) and resulted in potential underestimation of the difference between the two groups of schools.

²⁴ The 24 schools that participated in three years of that study (11 primary, 12 secondary and one P-10 school) were not a representative selection but were selected in that study because they were involved in reform.

²⁵ Of course those teachers were grouped in a small number of schools and it could be hypothesised would be more like each other than teachers in other schools.

teacher provided a more stable estimate of what happens in a classroom²⁶. Another would have been to use a multi-level analysis incorporating observation, teacher and schools as levels and compare trial and non-trial schools. The procedure adopted has consequences for the precision of estimates and whether apparent differences should be considered statistically significant. This is discussed further in the section concerned with the results of the study.

For the analysis the primary school classroom observations were treated as one group (i.e. not separately for each Year level) and the Year 8 classroom observations were treated as another group.

Results

Table 7 records the mean scale values for each dimension of productive pedagogies, level of schooling and year of observation. These values provide background for the discussion that follows²⁷. Analysis of whether differences in the reported values represented “real” differences that could be interpreted with confidence depend on taking account of the variation in scores within groups, the fact that the scale means were based on the same set of observations and the fact that there were multiple observations per teacher.

Table 7 Mean Scale values for Productive Pedagogies Dimensions

	Dimensions of Productive Pedagogies			
	Intellectual Quality	Connectedness	Supportive Classroom Environment	Recognition of Difference
All Observations				
2001	2.56 (0.82)	3.38 (1.15)	3.37 (0.56)	2.52 (0.70)
2002	2.31 (0.76)	2.64 (0.92)	2.95 (0.55)	2.04 (0.63)
2003	3.14 (1.07)	3.41 (1.11)	3.60 (0.74)	2.87 (1.03)
Year 5 Observations				
2001	2.43 (0.44)	3.10 (1.00)	3.30 (0.73)	2.38 (0.30)
2002	2.19 (0.67)	2.51 (0.86)	2.95 (0.57)	2.00 (0.55)
2003	3.22 (1.13)	3.43 (1.13)	3.77 (0.61)	3.16 (0.91)
Year 8 Observations				
2002	2.34 (0.73)	2.99 (0.88)	2.82 (0.59)	1.84 (0.60)
2003	2.97 (0.92)	3.34 (1.10)	3.23 (0.86)	2.22 (1.03)

Note: Standard deviations are shown in parentheses

Source: Data from Assessment and New Basics Branch and used in Research Paper No. 10.

The analyses for the longitudinal data in primary schools indicated that there was a drop in scores on all scales from 2001 to 2002 (that averaged 0.6 of a standard deviation) followed by a rise from 2002 to 2003 (that averaged 0.9 of a standard deviation) that resulted in a net gain (that averaged 0.35 of a standard deviation)²⁸. The analysis reported

²⁶ This would be the preferred approach in my view.

²⁷ These data were as recorded in a version of the research report accessed on 29 March 2004. The data are not reported in the final version of the research report but have been used as the basis for the graphical representations in Report Number 10.

²⁸ The net gains were greater for intellectual quality (0.6) than for the supportive environment and recognition of difference dimensions (0.4). There was almost no net change for connectedness.

in New Basics Research Paper Number 10 appeared to be based on observed lessons as the unit of analysis and indicated that the net gain was statistically significant for the intellectual quality, supportive environment and recognition of difference dimensions. When the number of teachers, rather than the number of observations, was used as the basis for computing standard errors these differences were not statistically significant even though the net gains were of moderate size²⁹.

The separate Year 5 and Year 8 analyses suggested that the gains were greater in Year 5 (average effect sizes from 2001 to 2003 of 0.8 and from 2002 to 2003 of 1.3) than Year 8 (average effect size from 2002 to 2003 of 0.5). In Year 5 the net gains were greatest for recognition of difference and intellectual quality whereas in Year 8 the gains were greatest for intellectual quality and supportive environment³⁰. The numbers at each of these Year levels are small and the gains are not significant when the numbers of teachers are used to estimate standard errors. If observed lessons were used as the basis for error estimation the gains on intellectual quality would show as significant at each Year level and on recognition of difference at Year 5.

The other observation that can be made from the data in Table 7 is the substantial increase in the variation (as reflected in the standard deviation) in Intellectual Quality and Recognition of Difference scores for Year 5 over the period from 2001 to 2003. This suggests that the net gain for some classrooms was substantial while others remained the same or perhaps declined. There was also an increase in the variation of recognition of difference and supportive environment scores for Year 8.

Interpretation

The overall conclusion from these studies is that from 2001 to 2003 there appears to have been an increase in the use of three of the productive pedagogy strategies (intellectual quality, supportive environment and recognition of difference) in trial schools. The magnitude of the apparent gains was greater for Intellectual Quality than for the other schools and was of moderate size. Although the statistical significance of the net gains remains uncertain the replication of the pattern across three separate scales provides some confidence that the gains might not be the result of chance fluctuation. To this extent there is support for the proposition that in trial schools there was an increase in the use of productive pedagogies during the course of the trial and especially in 2003. It is possible that changes observed in 2003 were associated with a more intense adoption of New Basics in trial schools.

Within the limitations of the size of the data set it appeared that the gains were greater in Year 5 than in Year 8. It also appeared that there was an increase in the variation among schools in several of the scales. This suggests that teachers implemented the productive pedagogy strategy to an increasingly varied extent. It would be interesting to know more

²⁹ The question of statistical significance involves taking account of the complexity in the data. The procedure used here is an approximation (that uses the average number of observations per teacher) and the preferred method that would use the teacher as the unit of analysis. The analysis using the teacher as the unit assumes that the multiple observations are intended to provide a more stable indication of what happens in that classroom.

³⁰ In Year 5 the effect sizes for the net gain from 2001 to 2003 were: 1.0, 0.3, 0.7 and 1.3 for the scales as listed. In Year 8 the effect sizes for the gain from 2002 to 2003 were 0.8, 0.4, 0.6 and 0.5 for the scales in the order listed.

about the conditions that resulted in greater use of those strategies but that would require a larger sample.

Productive pedagogies were also promoted as part of wider programs of professional development teachers across Queensland schools at the time the trial was being conducted. Any changes observed over time in trial schools might have also occurred in non-trial schools. The absence of comparative data from non-trial schools prevents an investigation of this possibility.

The Enacted Pedagogy Study

A second study concerned with changes in teaching and learning focused on the experiences reported by students in Years 5 and 8 using an on-line questionnaire called the Enacted Pedagogical Assessment (EPA) which was designed to assess student perceptions of the teaching in their classrooms.

Questionnaire

The questionnaire consisted of 51 items that measured 17 of the original 20 elements of productive pedagogies (it omits the elements deep knowledge and deep understanding as well as inclusivity) and through those the four dimensions of Productive Pedagogies. For each item there was either a positive or a negative statement to which students responded on a five-point scale to indicate their level of agreement. Students were invited to comment on their classroom experiences and supplied information such as date of birth, sex, year level and the name of their school.

From the item responses, indicator scores (referring to the elements of the framework) were computed as the average of the contributing item scores, and the dimension scores were the average of the indicator scores³¹. Reliabilities, as measured by coefficient alpha, ranged from 0.49 to 0.68, which are lower than the desirable minimum of 0.70 (Nunnally, 1978)³². The value of 0.49 for the intellectual quality scale is too low to be of value. The effect of low reliabilities would be to reduce the observed measures of an association between two variables and to distort patterns of results in various forms of multivariate analysis in ways that cannot be predicted.

Method

A two-stage sampling design was adopted with schools selected in the first stage and students selected in the second stage. Thirty-four of the 38 trial schools were invited to participate and all but two agreed to participate³³. Using a procedure based on a cluster analysis of selected demographic variables a set of 40 comparison schools was selected from the remaining Education Queensland schools. This is a sound way to establish a set of comparison schools. Valid data were received from 61 of the invited schools (32, trial

³¹ As noted previously an option would have been to weight the contributions from each item to the total score and to allow items to contribute to more than one scale.

³² Nunnally (1978) refers to this in the context of needing to adjust measures of association for unreliability rather than spend resources on improving reliability. It was in the context of aiming to achieve a higher reliability. A reliability of 0.7 means that half of the variance in observed scores is due to variance in the construct being measured and half is due to error. Multivariate methods of analysis usually demand higher levels of reliability.

³³ The four schools not included were three special schools and a school of distance education.

schools and 29 non-trial schools). Participating schools selected two classes from Year 5 or Year 8 (or from both if the school covered both Year levels)³⁴. The response rate was high (82 per cent of the designed sample responded), which means that there was very little possibility of bias in the sample.

Results number

In this discussion of the results (from New Basics Research Report Number 9) attention is focussed on the four dimensions of the EPA rather than the elements of each dimension. Table 8 summarises the mean scale scores (with standard errors) by Year and trial status for the four dimensions of productive pedagogies. It is important to caution against drawing conclusions from a simple comparison of means but the general pattern that is evident was supported by more sophisticated analyses.

Table 8 Means and Standard Errors for EPA Dimension Scores by Year Level and Trial Status

	Year 5		Year 8		Combined	
	Mean	SE	Mean	SE	Mean	SE
<u>Trial Schools</u>						
Intellectual Quality	3.37	0.02	3.20	0.02	3.29	0.01
Connectedness	3.61	0.02	3.25	0.03	3.43	0.02
Supportive Environment	3.76	0.02	3.34	0.03	3.55	0.02
Recognition of Difference	3.55	0.03	3.16	0.04	3.36	0.02
N	747		297		1044	
<u>Comparison Schools</u>						
Intellectual Quality	3.31	0.02	3.19	0.03	3.25	0.01
Connectedness	3.50	0.02	3.19	0.04	3.34	0.02
Supportive Environment	3.60	0.02	3.32	0.03	3.46	0.02
Recognition of Difference	3.43	0.03	3.05	0.05	3.24	0.03
N	673		403		1076	

Source: Information provided by Assessment and New Basics Branch in Version 1 of Research Paper 9

The first observation was that scores for Year 8 are lower than Year 5 in all four dimensions of the EPA in both trial and comparison schools. The difference was statistically significant. This is evident in other studies that rely on student reports. The second, and more salient, observation was that at Year 5 the average scores for comparison schools were statistically significantly lower than trial schools on three of the four EPA dimensions (but not intellectual quality). The differences between trial and non-trial schools at Year 8 were much smaller and not statistically significant.

In terms of elements within the major dimensions, it was worth focussing on two that were defined by five or more items. The element of the intellectual quality dimension with the greatest number of items (six) was higher order thinking. For that element, there was no difference between trial and comparison schools. From the supportive environment dimension, the element called engagement contained six items. The analysis showed significantly higher levels of engagement in trial schools than comparison schools at Year 5.

³⁴ There were procedures that randomly selected 60 students from schools with more than 60 students in the two classes and all students were included from schools with fewer than 60 students.

It was possible to use the standard errors and the numbers of respondents that are recorded to estimate the standard deviations. Effect sizes could be estimated as the difference in means divided by the standard deviation of the control (comparison) group. The effect size for the difference between trial and non-trial schools was greater in Year 5 than in Year 8. At Year 5 the effect sizes were 0.12, 0.21, 0.31 and 0.15 for intellectual quality, connectedness, supportive environment and recognition of difference³⁵. The corresponding values in Year 8 were 0.02, 0.06, 0.03 and 0.08 respectively.

Interpretation

The results for connectedness, supportive environment and recognition of difference, from the EPA study, are generally in the direction expected on the basis of the New Basics premises. However, the magnitudes of the differences between trial and comparison schools are small. The average effect size between trial and comparison groups for these three dimensions was 0.13 standard deviation units (but 0.22 in Year 5 and 0.06 in Year 8). By conventional criteria these are considered to be small effects. Another way of expressing this would be to say that the variations among trial schools, and comparison schools, were much greater than the difference between the two types of school. There was a larger difference between trial and comparison schools for supportive environment at Year 5. One factor that would diminish the observed differences between trial and non-trial schools was the unreliability of some scales in the instrument and especially the Intellectual Quality scale. The issue of the statistical significance of differences in a two-stage sample is complex. The two-stage design of the sample means that standard errors and tests of significance need to be considered either by including estimates of the design effect or by using multilevel analysis techniques. The structure of the data means that standard errors are likely to have been underestimated and differences of borderline significance are probably not significant.

Summary

The two studies concerned with teaching and learning practices provided complementary perspectives on the use of productive pedagogy strategies. One was from the perspective of observations of classroom practices. The other was from the perspective of students in classrooms. The observations of classroom practice indicated that in trial schools there were increases in the use of productive pedagogy (and consistently on Intellectual Quality) some of which were of moderate to large magnitude, even though there was some uncertainty regarding statistical significance. The increases were generally greater in Year 5 than Year 8. The increases in Year 5 were greatest in 2003 (in fact there was a decline from 2001 to 2002), which corresponds to the year in which the Rich Task assessment took place. The perspective from student responses to the EPA questionnaire indicated that student perceptions of productive pedagogy were a little greater in trial than non-trial schools (and possibly greater than that suggested by the instrument). Taking the two sets of data together there is support for the proposition that the use of productive pedagogies was enhanced as a result of participation in the trial even though it is not certain as to what extent the increase could be attributed to the trial.

³⁵ Estimated as the difference in means divided by the standard deviation of the comparison group. The standard deviation was estimated as $se \cdot \sqrt{n}$

ASSESSMENT

Assessment is a vital component of systems of schooling because it informs teachers, students and parents about progress and achievement. Assessments also represent manifestations of what is regarded as important in an education system and in the programs within the system. Studies in other countries (Earl et al, 2002) have indicated that assessment could be a powerful influence on the way a new program was implemented and the extent to which it was implemented. Despite this, assessment is sometimes resisted as a necessary evil rather than a valuable source of information. The assessment system developed and implemented in the New Basics trial was important to that trial and a set of research projects addressed the way it operated in practice. One project examined the extent to which teacher views of assessment may have changed during the course of the trial, another compared the key elements of the system with ideal types, a third examined the ratification process and a fourth examined the extent to which the assessment system captured outcomes that might not have been measured by other methods.

Teachers' Conceptions of Assessment

One of the research projects examined teachers' conceptions of assessment. It made use of a survey instrument that measured four dimensions of assessment: improvement, irrelevance, school accountability, and student accountability. It was expected that teachers in trial schools would have different conceptions of assessment than their peers in comparison schools because of the influence of the views of assessment contained in the New Basics material and the experience of assessing Rich Tasks.

Questionnaire

This was based on a subset of a longer questionnaire that had been used in other studies (Brown, 2002). The reliability coefficients for the first three of the scales were satisfactory at approximately 0.8 but that for student accountability was too low (0.5). Consequently it is possible to base inferences on results from the improvement, irrelevance and school accountability scales but not on the student accountability scales.

Method

The survey was administered as part of a larger questionnaire in November 2003. The sample included all 38 trial schools, all 21 Phase 2 schools, and 33 comparison schools chosen randomly from within those Education Queensland 'like' school categories that had Phase I schools. A total of 82 schools participated returning 1525 teacher responses (making approximately 19 teachers from each participating school). The survey methodology was sound. Approximately 90 per cent of all schools agreed to participate and the response rate from teachers in those schools was approximately 50 per cent. The low response rate from teachers means that there is a potential bias in the achieved sample that needs to be considered when results are interpreted.

Results

There was no significant difference in teachers' conceptions of assessment on these scales between trial and comparison schools. There were statistically significant differences between Queensland and New Zealand teachers (where the instrument had been developed) with Queensland teachers scoring higher on the purpose of assessment being to improve learning but lower on assessment being for school accountability. There were small, but statistically significant, differences between levels of schooling with improvement being more emphasised by teachers in primary schools and student accountability being more emphasised by teachers in secondary schools (although as noted previously the student accountability scale had a low reliability).

Interpretation

The study concerned with teachers' views of assessment indicated that Queensland teachers had a strong belief in the purpose of assessment being to improve individual student learning and a much less strong belief in the accountability purposes of assessment. There was no difference between trial and comparison schools in these emphases. Arguably, there are benefits in using assessment data to inform program development and planning. There was no evidence in these data of change having resulted from participation in the moderation process during that year.

Teacher Assessments of Rich Tasks

The assessment of Rich Tasks by teachers, and subsequently moderated, was an essential part of New Basics. Guides called grading masters supported those assessments. One of the important investigations in the research program concerned the way teachers assessed Rich Tasks using those grading masters. The study aimed to explore what teachers did when determining a grade and to identify the implicit rules that they used. Results from the study were intended to inform the development of teacher skills in standards-based assessment.

Background

Rich Tasks were described as complex and multi-faceted. They represented the culmination of a long period of work and drew on multiple skills. The approach to assessment asked teachers to judge the quality of student performance against several (typically three) specified features. It was possible to balance inconsistent ratings across features and assign an overall grade (a balanced score card) although some features could not be discounted. Coverage of the four categories of New Basics was required. The specified features were broad and substantive rather than narrowly specific. Teachers used a task-specific display incorporating descriptors (the grading master) to plot a performance for each of the features on a vertical axis (pole). They then made an on-balance judgment expressed as grade. This approach was different from those that use detailed rubrics, analytic marking and combination rules.

Source

Grading masters from the collections of student work submitted by trial schools for ratification at the end of 2003 were analysed. Additional grading masters were also gathered so as to ensure adequate coverage. In Year 3 there were five tasks and an

average of 188 forms from an average of 21 schools for the analysis. In Year 5 there were seven tasks with an average of 138 forms for each task and these were drawn from an average of 15 schools. In Year 9 there were eight tasks and an average of only 33 forms from an average of six schools. In total approximately 2,600 grading masters were analysed.

Method

A pattern analysis was conducted of annotated grading masters that represented six per cent of the total. Patterns of teacher classifications of student work across different poles were analysed and assigned as either simply mechanical, impression marking or trade-offs. Evidence was also collected from 12 of the 17 moderators involved in the trial. They recorded observations and reflections on the ways teachers used the grading masters to arrive at a grade. Moderators also recorded impressions of how teachers thought grading masters were to be used.

Results

Both the analysis of the grading masters and the moderators' comments suggested that this approach to assessment was a challenge for teachers. Approximately eight per cent of results were consistent with the intended approach in making use of trade-offs across features. A further 15 per cent were regarded as possibly consistent with the approach. Up to 77 per cent of cases could be interpreted as either impression marking followed by completion of the grading master (46 per cent) or as essentially analytic grading using mechanical rules to determine the final grade (31 per cent). It appeared that the rules about "non-discountable" poles were not operating as intended. The data were consistent with the notion that the tasks were multi-faceted. The small amount of data for Year 9 may indicate that there were difficulties in making assessments at that level.

Interpretation

The collection of grading masters was voluntary and might have been biased towards those who would implement the assessment model as intended. The study was intended to gain a general understanding of teachers' decision-making about grades but it did not investigate the detail of those decisions. It indicates that there is a need for ongoing professional development in standards referenced assessment.

The Rich Task Assessment System

Rich Tasks were a key element of New Basics and potentially they provided a basis for accountability to a range of interested parties. Moderated assessment of the Rich Tasks was proposed in the New Basics Technical Paper (Education Queensland, 2000). An investigation of the assessment systems examined its cost-effectiveness, elements that could be omitted, and the extent to which it generated comparability of standards.

Sources

Two sources of data were used for this project. One was the observation and reflection of those involved in the assessment system in various roles. The other was the collection of data on grades and changes during the moderation process.

Method

The Rich Task assessment system as implemented in the trial was compared with the model envisaged in the technical paper and with other moderation models. Key elements of the system were compared with the five elements of an effective assessment system and with literature about good assessment tasks. The first four rounds of the moderation strategy were analysed in relation to key aspects in the literature. Ratification forms from 2003 were analysed to ascertain whether there was an impact on levels of agreement and teacher-moderator interactions were analysed. Finally there was an analysis of costs.

Results

From the analysis of what happened in relation to other systems it was concluded that the system of moderated school-based assessment possessed the key elements of an effective system: articulated curriculum intent, curriculum plans or programs, evidence of learning, assessment standards and validation of judgments of student achievement. It was also argued that some elements needed to be strengthened such as curriculum planning from the Rich Tasks to fields of knowledge and skills. It was reported (in New Basics Research Report Number 23) that an assessment regime had been implemented with appropriate infrastructure and resources and that it was widely accepted.

There was variation in teachers' understandings of the grading process and moderation system. Most had an awareness of the need for grades to be based on the evidence from the submitted material, the elements of the grading masters and how those elements were used, and the role of moderation, and that comparability of standards required common assessment criteria.

The process of ratification had effects on grades that were different at the different juncture years and which varied across tasks within each year. Table 9 records these effects. At Years 3 and 6 approximately 60 per cent of grades were unchanged and between 25 and 30 per cent were regraded downwards. Fewer than 10 per cent were regraded upwards. At Year 9 only 28 per cent of grades were unchanged with 37 per cent being regraded downwards and 20 per cent being regraded upwards. In 15 per cent of cases the range was reduced.

Table 9 Classification of Level of Agreement in Grade Ratification

Impact type	Percentage of available forms		
	Year 3	Year 6	Year 9
Agreement	61	56	28
Some upward movement	5	8	19
General upward movement	1	1	1
Some downward movement	16	18	20
General downward movement	9	11	17
Top down, bottom up, middle unchanged	7	6	15
Top up, bottom down, middle unchanged	0	0	0
Number of forms	123	140	75
Median number of schools	24	22	10

Although grading masters provided guidance on the use of the desirable features when judging the standard of student work there was confusion about standards. Many schools and teachers were well disposed towards continuing their involvement and developing their expertise. The moderation process provided professional development for teachers in assessment and in content knowledge. It was suggested that teachers' resistance to assessment decreased and that they came to view assessment as part of learning (Earl, 2004).

Interpretation

From the analysis of the ratification process it appeared that the assessment system operated better at Years 3 and 6 than at Year 9, although even at those levels two out of every five assessments were amended. This suggested that professional development needed to be focused at junior secondary school or, possibly, that the grading masters at that level needed to be reviewed. The comparison of the assessment system with features of an ideal system suggested that it contained most important features. However, it would be useful to have a review by an expert not so closely connected with the present system. The moderated assessment system is important because there is no other systematic review of the standards of student work in Queensland below Years 11 and 12 (other than the statewide testing of basic skills). The data from the analysis of ratification processes suggested that teachers were tending to award higher grades than moderators believe are warranted.

It appeared that a moderated assessment system could provide feedback to teachers, schools and others about intended and achieved standards across the system. It could become an important element of the education system more generally as well as to New Basics. From other data it appeared that the moderated assessment system was important in promoting change in the nature of student work and teaching practices. A moderated assessment system provides important feedback but it does not involve consequences for individuals in the sense that an assessment used for entry to a course of professional study might. In that sense the moderated assessment system provides information of consequence but is not high stakes for individuals.

Profiles of Results on Rich Task Assessment in 2003

This project investigated the assessment characteristics of the Rich Tasks by investigating the patterns of results in 2003 in terms of the extent to which schools and students completed the tasks, the range of grades and the extent to which tasks were distinct or shared common underlying dimensions of achievement.

Source

Data in the form of grades on Rich Tasks from 4,656 students in 2003 (1201 in Year 3, 1164 in Year 6 and 2291 in Year 9) were analysed. Part of the analysis involved scores on the statewide Aspects of Literacy and Numeracy tests. Grades were either in four categories (A, B, C, C-) or six (A, B/A, B, C/B, C, C-) with additional codes of U for 'ungraded' and N for 'not given opportunity'.

Method

The analyses examined coverage, or compliance, by schools and students, the distribution of grades awarded for Rich Tasks, the dimensionality in the Rich Tasks using a principal components analysis, and the extent to which grades were related to literacy and numeracy skills.

Results

The evidence of the compliance rates of having no opportunity to attempt a Rich Task, or not completing a Rich Task, at both school and student level indicated variation. On average, across all Rich Tasks, 21 per cent of students had no opportunity to undertake a particular Rich Task. This ranged from 9.5 per cent to 65.5 per cent and was generally higher in Year 6 than either Year 9 or Year 3. In terms of non-completion of Rich Tasks the figures were lower averaging 11 per cent. This figure was higher in Year 9 (21 per cent) than Year 6 (10 per cent) or Year 3 (4 per cent). The data suggested some non-compliance with the requirement that all students complete all Rich Tasks that varied between Year levels and tasks. In general students in Year 6 were less likely to have had an opportunity to complete the tasks that were most different from existing practices.

The analysis of the distribution of grades showed no evidence of a ceiling effect (a high percentage of top grades). The median percentage receiving an A grade was three per cent with 16 per cent receiving “either A or the next highest grade”. A median proportion of 45 per cent were awarded a C grade, which carried the descriptor of commendable work (the minimum for any Rich task was 33 per cent). The C- grade was not used in large numbers (being typically less than 20 per cent) except for one task at Year 6.

The results of a principal components analysis at each Year level indicated that at each Year level there was one principal component (based on an eigen value greater than one³⁶) for each set of Rich Tasks. For each of Years 3, 6 and 9, the first principle component accounted for 51, 50 and 43 per cent respectively of the variance in Rich Task grades. This suggests that around half of the variation in grades was associated with this common component, which could be interpreted as richness. The remaining half of the variation was associated with unique characteristics of the task³⁷. There were some clusters apparent in Year 3. In Year 3 there were common elements in tasks 4 and 5 (Read and Talk about Stories and Historical and Social Aspects of a Craft) and tasks 1 and 2 (Webpage Design and Multimedia Presentation).

An analysis of Rich Task grades was conducted in combination with the scores for the same individuals on the Aspects of Literacy & Numeracy tests from the appropriate prior year. One limitation of the analysis was that the proportions that could be matched were low, ranging from 65 per cent in Year 9 to 78 per cent in Year 3. Regression analyses were conducted with the grades for each Rich Task as dependent variables and the test scores as independent variables³⁸. The proportion of the variance in grades accounted for

³⁶ Also the result of using the scree plot justified one principal component being extracted.

³⁷ This is based on an interpretation of the unrotated components. Rotation would emphasise interpretation in terms of unique factors.

³⁸ Strictly this form of analysis requires linearity in the relationship with grades and interval data. Even though these requirements could not be met with these data the results provide a good indication of the relationship. An alternative could have been to use a multinomial logistic regression and to

by test scores was around 13 per cent in Years 3 and 6 but only nine per cent in Year 9³⁹. The range was from three to 18 per cent. A combination of literacy and numeracy scores was a predictor of the Rich Task grade for 12 of the 20 tasks, literacy alone in six tasks and numeracy alone in just two tasks. Overall, the results indicated that literacy and numeracy scores are associated with rich task grades but were not strong predictors of those grades.

Interpretation

The analysis of grades for Rich Tasks in 2003 was especially valuable for understanding how the assessment system was operating. There was less than full completion of all tasks in trial schools. On average a little more than one in ten “student-tasks” were not completed but for Year 9 the figure was one in five. The reports indicate that this was partly associated with students transferring between schools but in high schools was for a range of other reasons associated with the tasks extending over three years. However, it was evident that the range of grades available was being utilized. It was also evident that Rich Task grades at any level represented an underlying expertise that was common and elements uniquely associated with each task. It would be of significant interest to explore further the nature of the expertise that was common. Finally, the analyses revealed that although the Rich Task grades were associated partly with literacy and numeracy skills, they embodied wider capacities.

Summary

The set of projects concerned with assessment provided key perspectives on an important aspect of New Basics. Teachers had an orientation to assessment as a tool for improving student learning and less towards using assessment to evaluate school programs and approaches. Perhaps a greater orientation to those other purposes of assessment would encourage the development of Rich Tasks in the school setting. Although teachers in primary schools were able to use the grading masters competently to assign grades, more development of expertise was needed. The moderated assessment system contained the essential features of good assessment systems but it operated better in primary than in secondary schools. It seemed important to have found that there was a common element, in addition to unique features, in Rich Tasks and it remains important to understand better the nature of that common element. The Rich Task assessment was measuring expertise and capability that was wider than literacy and numeracy skills.

incorporate a multilevel design but the results would have been more complex to interpret. The regression analysis performed is sufficiently robust for the purpose.

³⁹ This corresponds to multiple correlation coefficients of between 0.30 and 0.36.

STUDENT ACHIEVEMENT

New Basics was intended to increase the intellectual quality of students' work. In addition students who had been intellectually engaged would be expected to show improved higher order thinking skills and more general skills of problem solving (Education Queensland, 2003a: 12-14). In other words, did enhanced richness in student work transfer to new and unknown situations. It was also intended that student performance in basic literacy and numeracy "would, at worst, not drop because of the experience of the New Basics" (Education Queensland, 2003 a: 13).

There were three studies in the research program concerned with different aspects of student performance: a study of performance in relation to student and school characteristics on the International Schools Assessment; an analysis of patterns and trends in the state-wide tests of literacy and numeracy; and an analysis of performance on a computer-based assessment of problem solving. Each contributed to an understanding of student achievement in different ways.

The International Schools Assessment

The International Schools Assessment (ISA) was based on the reading and mathematical literacy frameworks of the OECD Programme for International Student Assessment (PISA), which is a survey of the knowledge and skills of 15-year-olds in OECD countries. It measures skills in reading literacy, mathematical literacy and writing for students in Years 3, 5, 7 and 10 in international schools. Reading and mathematical literacy, in these frameworks, involved higher-level skills than were normally associated with basic literacy and numeracy and which could be seen as consistent with the higher-order skills valued in the New Basics⁴⁰. There was also an assessment of writing as part of the assessment. The process of using the ISA provided individual level data about student background and thus enabled the influence of social and personal background characteristics on achievement and growth in achievement to be investigated in trial and non-trial schools.

Assessment

These frameworks described reading literacy as "understanding, using, and reflecting on written texts, in order to achieve one's goals, to develop one's knowledge and potential, and to participate in society" (OECD, 1999: 20). The descriptions added that the "assessment of reading literacy requires students to perform a range of tasks with different kinds of text. The tasks range from retrieving specific information to demonstrating a broad understanding and interpreting text and reflecting on its content and features." The material included both continuous texts (prose organised in sentences and paragraphs, in a range of prose forms) and non-continuous texts (such as in lists, forms, graphs or diagrams). The tasks require students to "demonstrate their proficiency in retrieving information, forming a broad general understanding of the text, interpreting

⁴⁰ The higher order skills embodied in the ISA are set within the domains of reading, mathematics and writing rather than in the domains that are the organisers for New Basics.

it, reflecting on the content and form of texts in relation to their own knowledge of the world, and arguing their own point of view”.

Mathematical literacy was defined in corresponding frameworks as “an individual’s capacity to identify and understand the role that mathematics plays in the world, to make well-founded mathematical judgements and to engage in mathematics, in ways that meet the needs of that individual’s current and future life as a constructive, concerned and reflective citizen”. Those frameworks involved dimensions of content (change, change and growth, space and shape, reasoning, uncertainty and dependency relationships), process (mathematical language, modelling and problem-solving skills), and mathematical thinking, generalisation and insight.

Writing was assessed through pieces of narrative and expository writing that were intended to cover the expression of ideas and organisation of content, language structure (sentence construction, grammar, vocabulary) and spelling.

For each domain a common scale was used for all Years so that a score of 450 for a Year 3 student indicated the same level of proficiency as a scale score of 450 for a Year 10 student and the same level of proficiency from one year to the next. Item-response theory (IRT) and link items (items that are common to different tests) were used to calibrate scores to a common scale⁴¹. For this reason it was possible to monitor individual progress over time and measure growth in learning (ACER 2002). Because of the differences between the year levels targeted by the ISA and the New Basics juncture years, adapted tests were used in Years 6 and 9. Students who sat for the tests also completed a questionnaire about their age, gender, and family background.

Methods

The ISA was administered as a component of the research program in March 2003 and December 2003. Fifty-eight Education Queensland schools (34 trial and 24 like schools) including 7,627 students in Years 3, 6 and 9 provided data in March for the main study⁴². Fifty-six of these schools also provided data from 8,249 students in December^{43,44}. In each participating school, all students at each Year level were requested to participate.

Item Response Theory (IRT) was used to calibrate the March and December results on the same scale. The change in ISA scores for students from trial and “like” non-trial schools from March to December 2003 was computed and reported for each year level and the distributions of ISA results for students from trial and “like” non-trial schools in December 2003 for each year level were compared. Finally, and most importantly, multilevel modelling of results was undertaken using the variables (at the student level)

⁴¹ Scores on mathematics and reading were the calibrated ISA scales. Scores for writing were the sums of the raw scores on two separate writing tests.

⁴² In addition 27 non-government schools (3,536 students) participated in the trial in March of which 26 (with 3,737 students) participated in December. They do not form part of this analysis.

⁴³ The increase in the number of students even though there were two fewer schools probably reflects greater compliance, in December than in March, with the request for all students at each Year level to participate. Measures of growth were calculated for only those students with both achievement scores (December and March). There was some controversy about the tests in March.

⁴⁴ An additional 23 schools (including 3 new Basics Phase 2 schools) including 1643 students provided data in December as part of a control to check on any practice effects. They do not form part of this analysis.

of gender, Aboriginality (Aboriginal or Torres Strait Islander), socio-economic index based on father's job, and results in the statewide literacy and numeracy tests, and (at the school level) of remoteness and index of disadvantage⁴⁵. Using analyses that incorporated this range of variables meant that issues of the extent to which "like" schools resembled the trial schools were far less important because data on student background and social context had been included in the multivariate analyses.

Results

The reading, mathematics and writing proficiency of students improved over the period from March to December in each of the year levels and in both trial and non-trial schools. The change over the period was greater in Year 3 than Year 6 and greater in Year 6 than Year 9. Moreover, the difference between average scores at Year 3 and Year 6 was greater than the difference in average scores between Year 6 and Year 9. This was consistent with other studies that had examined the profiles of growth over the course of schooling (Rowe & Hill, 1996).

The initial analysis of the means recorded in Table 10 suggested, and the multilevel analyses established, that there was no significant difference between the growth of students in trial and non-trial schools in any domain in any year level. Nor was there any significant difference between trial and non-trial schools in the end-of-year achievement after allowing for the influence of other factors in the analysis.

Table 10 Mean Scores on ISA Tests by Time, Year Level and Trial Status

	March				December			
	Trial		Non-Trial		Trial		Non-Trial	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Year 3								
Mathematics	279.9	75.5	263.6	69.1	319.9	78.3	318.5	66.4
Reading	247.3	111.2	234.9	102.4	322.8	111.9	327.6	103.1
Writing	16.8	4.9	16.7	4.6	21.4	4.9	21.6	4.2
Year 6								
Mathematics	414.4	82.2	417.8	76.2	476.8	75.8	477.4	68.8
Reading	402.2	105.2	410.7	98.6	452.1	114.0	450.3	103.2
Writing	26.9	6.5	26.1	5.9	28.8	5.9	28.9	5.3
Year 9								
Mathematics	537.6	72.2	524.0	75.3	574.3	66.2	575.1	64.9
Reading	497.0	96.1	484.8	98.0	516.7	101.0	506.8	98.1
Writing	35.4	7.4	34.5	6.9	35.9	6.7	35.0	6.1

For Year 3 $N(\text{trial}) = 919$, $N(\text{non}) = 248$;

For Year 6 $N(\text{trial}) = 933$, $N(\text{non}) = 584$;

For Year 9 $N(\text{trial}) = 1735$, $N(\text{non}) = 1685$.

Source: Analysis conducted by Assessment and New Basics Branch

⁴⁵ Multilevel modelling was used because the data had a hierarchical structure involving students within schools and to allow for the influence of associated factors.

Based on the pattern of means it appeared that the difference, between trial and non-trial schools, in the change from March to December at Year 3 in mathematics was -15 units (or -0.21 standard deviations) and in reading was -17 units (or -0.17 standard deviation units). In writing at Year 3 the corresponding difference was -0.06 standard deviations. The corresponding changes at Year 5 were 0.04 , 0.10 and -0.15 standard deviation units for mathematics, reading and writing respectively. At Year 9 the corresponding changes from March to December were -0.21 , -0.02 and 0.0 standard deviation units for mathematics, reading and writing respectively. The average effect size was -0.08 standard deviation units. These measures of change do not take account of differences in the characteristics of students that may have been associated with the change. None of these measures of change was found to be statistically significant after using appropriate multilevel methods of analysis that made allowance for other associations among the variables.

There were other results from the multilevel analysis that were also salient to the New Basics trial. First, there was no difference between trial and non-trial schools in the effect of individual-level socioeconomic status, or of school-level socioeconomic context, on growth over the year in any of the domains. Second, Aboriginal and Torres Strait Islander students showed similar rates of growth as other students except in Year 6 reading in which case they showed higher rates of growth. Third, girls and boys showed similar development in reading and mathematics at each year level but girls showed greater development in writing skills⁴⁶.

Interpretation

The data gathered through the ISA provided a powerful basis for analysing some of the potential impacts of New Basics because they provided measures of outcomes at two times (March and December) and because they contained information about student background and school context that could be linked with the achievement measures. In addition because nearly all the students at each year level in the sampled schools provided data (although there was slightly higher participation in December than March) there was little systematic selection bias.

The general conclusion about New Basics was mixed. There was no evidence of enhanced achievement in these domains but neither was there any evidence of generally reduced performance⁴⁷. If students in New Basics trial schools were learning additional skills it was not to the detriment of their proficiency in these relatively complex aspects of reading, mathematics and writing.

In relation to the equity purposes of New Basics there was no evidence of greater equity being developed in relation to student socioeconomic background or in relation to the socioeconomic level of the area served by the school. New Basics had the intention of promoting greater equity in schooling. On this evidence it would appear that this had not happened during the period between the two administrations of the tests. On the other hand there was a view that curricular reforms such as this draw on community as well as school resources and could exacerbate differences related to social backgrounds and contexts. That did not appear to have happened.

⁴⁶ This result should be interpreted with a little caution because the writing measure was not on a calibrated ISA scale.

⁴⁷ Although in the unadjusted effects there was a lower rate of growth in reading and mathematical literacy at Year 3 in the trial schools.

There was evidence of enhanced growth in reading literacy for Aboriginal and Torres Strait Islander students in Year 6. This is a finding that deserves to be followed through as to why this was evident in Year 6 and not at other levels and why it was evident in reading literacy and not in mathematical literacy. It is possible that it might relate to particular tasks undertaken in and before Year 6. The suggestion of enhanced growth in writing literacy for girls is something that exacerbates a difference rather than reduces it but may reflect the greater opportunity for extended writing being available for students with those interests to extend their skills.

Statewide Literacy and Numeracy Assessments

Although the implementation of New Basics was intended primarily to enhance a wide range of skills related to the intellectual quality, it was also intended that there should be no decline in basic literacy and numeracy skills. There was a study of changes in the performance of students in trial schools on statewide literacy and numeracy tests over the period of the trial.

Assessment

The Queensland Studies Authority routinely administers statewide tests of literacy and numeracy in Years 3, 5 and 7. Those tests focused on the percentage of students achieving the benchmark level. This focus results in items that provide the greatest possible precision at the benchmark level and consequently the tests do not give as much emphasis to higher-level skills. There was a common scale within each domain with means defined as 500 for Year 3, 600 for Year 5 and 700 for Year 7. The absence of equating across Year levels meant that the potential for assessing growth was limited.

Method

The study used the data over the period from 2000 to 2003 inclusive to monitor trends in trial schools and all other non-trial schools in the government school system. This represented a different comparison group from the “like” schools and “comparison” schools used in other studies in the research program. It was not possible to track individuals from Year 3 to Year 5, or from Year 5 to Year 7, because the identification data could not be matched in a sufficient number of instances. Hence the analyses focused on trends within Year levels over time. It would have been a more powerful form of analysis to track individual students and apply similar methods of analysis to those used for the ISA.

The focus in the analysis was on the difference in the median scores for trial and non-trial schools. In other words the median for trial schools was subtracted from the median for non-trial schools and that “gap” index was studied over time. This “gap” index fluctuated from year to year rather than following a monotonic trend. The instability in the index is associated with the fact that it represents the small difference between large values so errors are compounded.

Results

The results showed that there was an initial gap in the performance of students in trial and non-trial schools of between 10 and 25 points in favour of students in non-trial schools (with 50 points intended to represent one year’s average growth). This indicated

that the schools in the trial schools were not a random, or even a representative, selection of government schools. Over the course of the trial the gap between trial and non-trial schools reduced by an average of approximately four points. The reduction appeared to be about five points in literacy and nearly two points in numeracy but there was considerable variation between Year levels. The only case in which the gap widened was for numeracy in Year 3. Relevant data are recorded in Table 11.

Table 11 Reduction of the Gap in Median Scores for Trial and Non-Trial Schools between 2001 and 2003

	Literacy Gap	Numeracy Gap
Year 3	6.3	-9.0
Year 5	4.7	12.3
Year 7	4.2	2.1

Source: Data computed by Assessment and New Basics Branch for Version 1 of Research Report 6.

Note: A positive value for this index indicates that performance in trial schools improved relative to non-trial schools and a negative value indicates that performance in trial schools worsened relative to non-trial schools.

It is difficult to estimate effect sizes from the statistics reported in New Basics Research Report Number 6, but given the distributional properties of the tests a five point change (the average for literacy) would correspond to an effect size of approximately 0.07 for literacy and a two point change (the average for numeracy) would correspond to an effect size of 0.02⁴⁸.

Interpretation

Interpretation of these trends needs to allow for the instability of the gap index. There is evidence of a small reduction in the gap between trial and non-trial schools and no evidence that trial schools deteriorated over the course of the trial in relation to the other state schools on the literacy and numeracy tests.

World Class Tests

There was a study that set out to investigate changes over time in the problem-solving performance of students in trial schools and compare the problem-solving performance of students in trial and non-trial schools.

Assessment

The study made use of the World Class Tests (WCT) in problem solving designed for 9-year-old and 13-year-old students and targeted the top 10 per cent of the ability range in problem solving⁴⁹. There was a paper-based and a computer-based component for each

⁴⁸ It is possible to interpret the difference in medians in relation to the standard deviations or the inter-quartile range. For literacy the standard deviations were 76, 69 and 64 units at Years 3, 5 and 7 with the inter-quartile ranges being 101, 89 and 78 units respectively. The corresponding standard deviations for number were 104, 97 and 101 and the inter-quartile ranges were 132, 120 and 111.

⁴⁹ The tests were developed by the Qualifications and Curriculum Authority (QCA) in London, and administered through the Assessment & Qualifications Alliance (AQA).

test. The test had face validity in that it was concerned with problem solving set in the context of rich stimulus material. However, because the difficulty was high, the distribution of scores in the Queensland context meant that it was limited in providing reliable measures of growth. This difficulty was compounded by the absence of appropriate equating data between test forms. Marking of completed tests was the responsibility of the agency that administered the process. Results were reported in four grades: distinction, merit, pass, and below pass.

Method

After trials in 2001⁵⁰ the tests were administered to Year 4 and Year 8 students in trial and like non-trial in April and July of 2002. Schools were asked to select the top 10 per cent of students to take part in the tests but there was little control over the selection process. In March 2003 the tests were administered to trial and non-trial schools again (348 Year 4 students from 35 schools participated and 427 Year 8 students from 22 schools participated)⁵¹. There were technical problems in the administration of the computer-based component of the tests that disrupted assessment procedures significantly.

Issues of sampling represented a threat to the validity of the exercise. First, not all trial schools participated and the achieved sample of non-trial schools was less than intended. Second, schools made the selection of the top 10 per cent of students thus limiting the generalisations that could be inferred from any analysis.

Results

Most students from both trial and non-trial schools received a “below pass” grade on the WCT. There was no reportable difference in results between trial and non-trial schools among either Year 4 or Year 8 students in 2003. In Year 4, 13 per cent of students from trial schools and the same percentage from non-trial schools received a pass grade or above. In Year 8 27 per cent of students from trial schools received a pass grade or above compared to 41 per cent of students from non-trial schools. For Year 8 students none of the variables (numeracy, literacy, gender, and school index of disadvantage) were related to performance. For Year 8 students numeracy results (from Year 7 in 2002) were associated with performance but only explained 25 per cent of the variance.

Interpretation

The nature of the selection of participating students and the technical difficulties in test administration meant that it was difficult to draw any conclusions from the study.

Summary

Results from the analysis of data from statewide literacy and numeracy tests indicated that there was no deterioration in performance on those basic skills associated with

⁵⁰ Schools (including some non-government schools) were invited to nominate an interest in participation.

⁵¹ There were 138 Year 4 students from 25 trial schools and 92 Year 4 students from 13 non-trial schools. In addition there were 210 Year 8 students from 10 trial schools and 217 Year 8 students from 12 non-trial schools. In addition some non-government schools participated but these were not included in the research project.

participation in the New Basics trial. There was some evidence of a very small reduction in the gap between the median scores in trial and non-trial schools which would be encouraging if it is sustained over future years. The International Schools Assessment (ISA) was concerned with higher order skills in the domains of reading literacy, mathematical literacy and writing literacy. These represent some of the areas to which it might be expected that the experience of richer quality work might transfer, either because of the skills developed or because of higher levels of engagement. The results indicated no significant difference between trial and non-trial schools in growth in scores on the ISA domains after allowance for other characteristics of students.

There was no differential impact between trial and non-trial schools of individual socioeconomic background on intellectual growth. Nor was there an effect of school-level socioeconomic context. Being in the New Basics trial neither reduced nor widened these equity differences. In general there was no effect of the trial on the differences between Aboriginal and Torres Strait Islander students and other students. However, in reading literacy at Year 6 there appeared to be enhanced growth for Aboriginal and Torres Strait Islander students. Although this is just one effect it does prompt the possibility of exploring further whether this effect endures in future cohorts and what might have been the reasons for the shift. There was also evidence of enhanced writing performance for girls compared to boys. Although this does not represent any narrowing of an existing gap, and in fact is the reverse of that, it does prompt questions about possible causes.

SCHOOL ORGANISATION AND STRUCTURE

The New Basics framework was concerned with structural and organisational change in schools in two ways. On one hand it recognised that implementation of New Basics required, either at the outset or introduced concomitantly, organisational capacities and arrangements that would facilitate its operation. On the other hand it envisaged that the process of working with New Basics would itself generate organisational change. The concept of reciprocal causality is common in social sciences and especially in relation to the impact of organisation and structure⁵². Some aspects of organisation and structure, such as the concept of the professional learning community are abstract and evidenced by manifestations such as how people interact with each other and with their tasks. Others are more concrete such as the administrative framework of a school, the timetable or the physical organisation of classrooms and other resources. There were several elements concerned with organisation and structure in the research program. One element was based on surveys of teachers and aimed to investigate the extent to which professional learning communities operated in schools. Others were based on detailed analyses of interviews with key stakeholders in schools and aimed to explore the ways in which changing organisations and structures facilitated the implementation of New Basics.

The Teacher Surveys

In literature associated with New Basics there was a proposition that professional learning communities were both necessary for pedagogical change and would be enhanced as a result of participation in New Basics. This grew out of research from the Centre on Organization and Restructuring of Schools (CORS), and built into the QSRLS, that suggested that organizational context was a key factor in teacher promotion of student learning. There were activities in the New Basics trial intended to enhance and support the teacher professional community. Trial schools were exempted from some curriculum, assessment and reporting obligations; they were given special assistance and resourcing, and funding for teacher release so that teachers could work together, with the school's critical friend or with staff from the Assessment and New Basics Branch. Professional development on productive pedagogies, Rich Tasks, grading and moderation, was also provided. Accordingly there was an investigation of whether there were stronger teacher professional communities in trial schools with the supporting elements of teacher empowerment and human and social resources.

Instrument

The instrument used for assessing the strength of the teacher professional community was based on one established by CORS work and used in the QSRLS (The University of Queensland, 2001; Louis, Marks & Kruse, 1996). The instrument assessed three aspects of school organisational context: professional community, teacher empowerment, and human and social resources. Each of these scales had subscales and items had five-point response categories or were open-ended. The reliabilities of scale scores for teacher

⁵² In this Chapter the terms have been used loosely and to some extent interchangeably. Structure can be taken to refer to established frameworks and arrangements whereas organisation has been taken to refer to how processes occur within structures.

empowerment and social resources were satisfactory (0.85, 0.79) and the reliability for professional community was barely adequate ((0.67). Two subscales (collective activity and deprivatised practice) had less than satisfactory reliabilities (0.63 and 0.66). Most of the analyses focussed on subscales rather than the three indices. The effect of lower reliabilities on comparisons of means is to underestimate differences that may really exist.

Method

The instrument was administered as part of a larger questionnaire in November 2002 and 2003. In 2002 all 38 trial schools and a set of 54 comparison schools (chosen randomly from the EQ 'like school' categories that had trial schools) were invited to participate. In 2003 all 38 trial schools, all 21 Phase II schools, and 33 of the 34 comparison schools that had returned at least one questionnaire in 2002 were invited to participate. There were 868 responses from 72 schools in 2002, and 1525 responses from 82 schools in 2003. The teacher response rates were 37 per cent in 2002 and 53 per cent in 2003. Low response rates were a potential threat to the internal validity of the study and the research report notes that the change in response rates may have been a source of the general rise in scores from 2002 to 2003.

There was skewness in the distribution of responses and a transformation was applied to item responses to reduce the effects of this. Subscale scores were calculated as the sum of transformed (to reduce the effect of skewness) but unstandardised component items. The subscales were then standardised. This procedure followed that used by Louis, Marks and Kruse (1996).

The structure of the achieved sample involved teachers within schools. Estimates of precision and the significance of differences using simple random sample procedures may possibly overstate precision and give a false appearance of significant differences.

Results

Trial schools were higher than non-trial schools in the four sub-scales of Collective Activity (Teacher Professional Community); Shared Decision Making (Teacher Empowerment); and Supportive Leadership and Openness to Innovation (Human & Social Resources). Between 2002 and 2003 there was an increase on eight of the sub-scales but no relative change between trial and non-trial schools on most of those scales. There was a relative change on the Professional Development over that year with a decline for non-trial schools and no change for trial schools. However, that relative change was of borderline significance ($p=0.03$) and may have not been evident if allowance was made for characteristics of the sample. Regardless of the issue of significance the effect was a small one and reflects the greater role of professional development in trial schools.

Interpretation

If it is assumed that there was no initial difference between trial and comparison schools then the evidence is consistent with trial schools having experienced a relative increase in collective activity, supportive leadership, and openness to innovation. Overall the results suggested that trial schools had higher measures of teacher professional community than comparison schools but that may have been a pre-existing state rather than something

that developed during the trial. Organisational capacity was one of the factors considered when trial schools were selected.

The Thirteen School Study

New Basics involved approaches that drew upon several disciplines and an orientation to problems in the world of work and everyday life. It was recognised in planning documents that its implementation would require a supportive organisation. This would involve consideration of structures, roles, timetables, budgeting and the use of space. The 13-school study documented the way schools responded to New Basics in terms of school structure, organisation and leadership. The 13-school study was also used to provide an analysis of links to external communities.

Instrument

Data were collected through an interview schedule that focused on areas relevant to the research questions and related to different levels of school organisation⁵³. The schedule was used as a guide to encourage participants to discuss what happened in their school and their perspective of changes rather than as a prescribed instrument. Issues covered in the interviews for all those interviews were:

- Accountabilities for implementing the New Basics and its implications for personnel
- Changes in formal roles in the school and the organisation of these adaptations
- Informal adaptations of roles as a result of the trial and their organisation
- Timetable influences (constrain or enable) on the implementation of the New Basics
- Informal adaptations to timetabling to accommodate New Basics implementation
- Adaptations (formal and informal) to the organisation of space in the school and within classrooms as a result of New Basics implementation?
- Views of resourcing issues with respect to the New Basics trial

Additional issues covered in the interviews with principals and deputy principals were:

- Adaptations to the personnel structure of the school to cater for implementation
- Amount of funding to the New Basics trial and the sources of that funding
- Allocation of funding across New Basics component areas

⁵³ In addition to focussing on questions concerning the central tenets of New Basics in relation to school organisational change, the interview schedule also incorporated different levels of organisational functioning as identified in research on schools as organisations: teachers' work, institutional linkages that bind schools into their societies, and less formal adaptations of structures.

Method

Thirteen of the trial schools were selected as case studies taking into account characteristics that were likely to be related to variations in factors that would facilitate the development of school professional learning communities: school size, location, school type, socio-economic demographics, cultural factors, participation in clusters, critical friend usage, and administrative approach. Six of the schools were in metropolitan locations, four were in regional locations, two were in rural locations, and one was in a remote area. There were three high schools, eight primary schools (one with P to 10 enrolments), a special school and a school of distance education. In each school principals, deputy principals, critical friends, teacher aides, and a minimum of two teachers were interviewed. These interviews were held early in the third year of the trial. All the interviews were transcribed so that they could be analysed systematically. A computer-based search of the transcripts for key terms was conducted first. Then data were classified on the basis of the ways school staff described these themes⁵⁴.

Results

The results of the analyses showed that almost all the schools understood that shared ownership of changes to curriculum and pedagogy was involved in the reform. The need for changes in structure varied among the schools, especially between primary and secondary schools. In some schools structural changes were aligned with dispersed leadership and shared ownership but in others this was not evident. Conditions prior to the trial, such as existing collaborative mechanisms, size and organisation influenced the alignment between structural change and the development of shared ownership. Some schools that did not have these conditions in place were able to introduce them during the trial.

All the primary schools had to make some structural changes especially in relation to collaborative curriculum planning across year level groupings. In large primary schools, this was more formal whereas in small schools that had been using multi-grade classes the changes were less formal. Changes in other functions such as timetabling, use of space, budget processes were also required. Changed budget and resource allocation mechanisms were important to the adoption of the New Basics model. Where only limited changes were made to roles and other organisational matters there was little alignment with the New Basics model of shared ownership.

In the secondary schools there was a greater emphasis on the structural changes that were necessary and in the two larger schools the creation of substructures (such as sub school arrangement and a middle school organisation) were seen as necessary to reduce specialisation by subject field and level and achieve collaborative curriculum planning. One of the other two schools (a small school) had built its collaboration around principles that had been established in another project. The other secondary school had attempted to introduce New Basics without structural changes. It maintained its structure based on traditional departments and was seen to have encountered obstacles in implementing the reform. In the three aligned schools there was change in curriculum organisation and timetables and in the non-aligned school the traditional forms of organisation had been retained. Initiatives in timetabling and subject offerings were important conditions for

⁵⁴ The authors of the paper provided a basis for identification of factors that differentiated 'aligned' and 'non-aligned' models of implementation.

shared ownership. Where these had been taken, the large high schools were able to align operational planning and administration with shared ownership. In the large high school where there was less alignment, staff did not adopt collaborative planning approaches and there was an absence of changes to subject offerings, timetabling adaptations and devolved budgeting.

Interpretation

This project included a rich and detailed array of data from a small number of schools. For that reason the findings based on the primary schools may be more robust than those based on the high schools. The gathering of evidence through interviews conducted at one point in the third year of the trial means that the information provided by respondents is based on reflections and retrospective interpretations of events during the trial. Nevertheless, the detailed information and careful analysis that includes cross-validation provided a depth of understanding of the process of implementing the New Basics trial.

From the results of the analysis it was clear that schools needed to implement changes in order to align planning and administration with New Basics. Alignment required the development of organisational frameworks that facilitated collaboration⁵⁵. This was most necessary in large schools although some small schools were able to achieve this without major changes. Staff needed to adapt to changes in their roles on an ongoing basis and to have a wider understanding of curriculum on a school-wide basis. Many aligned schools implemented mentoring and feedback systems to support the development of cooperation and negotiation.

In terms of organisational matters there was a need to adapt the use of space, timetables, subject offerings and budgeting process. Establishing spaces suited to collaboration was important. There was also a need to manage teacher release and to negotiate impacts associated with changes to subject offerings. Aligned schools instituted budget processes that were devolved, consultative and adaptable. These factors should inform future approaches to implementation of New Basics at the level of school or organisation.

The study indicated that the implementation of New Basics in high schools was more difficult than in primary schools. At least in part this may have been associated with the fact that the three-year span preceding Year 9 juncture extended back to the last year of primary school. With students in a high school coming from a wide range of primary schools this appeared to be a significant impediment to full implementation. In addition the organisational structures in high schools represented a potential barrier to implementation. Although this can be interpreted simply in terms of organisation another interpretation could involve the knowledge beliefs of high school teachers. In this sense the New Basics curriculum organisers may not have been seen as a way of connecting to the disciplines of high school teachers. Many of the Rich Tasks could be interpreted in terms of the knowledge and skills of those disciplines and that may be a way to facilitate uptake. There may also be an issue of how the work done in and through schools can be connected to the worlds of scholarship as well as to the immediate environment of students.

⁵⁵ Such as sub-school and middle school structures and team-based principles of organisation

Community Linkages

Data from the 13-school study were also used to investigate the links between schools and communities that developed during the New Basics trial. Planning documents for New Basics had envisaged wider roles for the community in supporting and facilitating the reform process and even incorporating the community part of the school professional learning community. In relation to the Rich Tasks it was asserted: “in developing and refining Rich Tasks, teachers and members of the community would be invited to attend panel meetings.” It was envisaged that communities would be involved in providing formative development feedback on the change. It also envisaged that the principle of connectedness would link student learning to what happened in their communities and that community specialists might be involved with programs.

Instrument and Method

The instrument was the open interview schedule used in the larger 13-school study documented above and the procedure for analysis was the same as reported for that study.

Results

There was some evidence of students going out into the community. Five schools reported students going out into community but four of these were part of specific requirements (oral histories or the improving wellbeing rich task). Community links were seen more often as problems in terms of finding appropriate expertise and logistics. More fundamental were issues of whether the trans-disciplinary focus of New Basics was understood in the community or matched the nature of work.

Most schools had not established new roles for other community members in schools and only one teacher reported consulting other professionals in developing lessons. Others questioned the availability of appropriate expertise. There was limited reference to the “life worlds of students” as a resource for pedagogy. Parents were involved in classrooms in traditional roles in seven of the schools. There was little evidence of communities providing formative feedback and usually with reference to parents rather than a wider sense of community. The evidence that was available involved established forms and transmitting information about New Basics.

Interpretation

Although the evidence was from a small number of schools the conclusion was that links with the wider community were evident in only a few schools that were able to build on pre-existing structures and here staff perceived the community to have a high level of relevant expertise.

The Role of the Critical Friend

Critical friends were appointed to schools to undertake school-level research; assist in writing school-specific documentation; assist school management in the process of reform; and provide advice to the central authority on implementation. One of the research papers documented the ways in which critical friends worked with schools and the system.

Sources

Data were gathered by means of questionnaire and discussion notes. Questionnaires were developed for school management team members, teachers, critical friends or staff in central office. These asked respondents to express a level of agreement with statements about the work of critical friends.

Method

Questionnaires were administered in the middle of 2002 and replies were received from approximately 30 respondents⁵⁶ in each of the four categories mentioned above. Data were also collected during a discussion day in June 2003 attended by all critical friends and principals of Phase I and II trial schools.

Results

Not surprisingly critical friends filled their roles in a range of ways that depended on the school and the individual orientation. They were most effective when contact with teachers was supported by the allocation of resources and time to provide opportunities to meet. Their influence was greatest in the areas of changes to teaching processes and assisting with writing school documents such as curriculum plans. The influence of the critical was facilitated by the location of the school being convenient for frequent contact, the allocation of funds and therefore the time available to support the work of the critical friend, an open structure in the school, good rapport with teachers, and the ability of the person to fill a number of functions.

Interpretation

Although the study was based on perceptions from a relatively small number of instances the results suggest that there could be more attention to the roles and expectations of critical friends in schools implementing innovations.

Summary

A number of themes concerned with school organisation and structure emerged from the research reports. In the background there was ambiguity as to whether a strong teacher professional community was a necessary condition for implementing New Basics, something that was developed through New Basics or a combination of these. From survey evidence it appeared that collective activity, supportive leadership, and openness to innovation increased during the trial. Overall trial schools had higher levels of teacher professional community than comparison schools but that may have been a state that predated the trial given that organisational capacity was a criterion for participation in the trial.

The case studies suggested that schools needed to make changes when they implemented New Basics, especially in large schools. Some of these changes involved the roles of teaching staff that was supported through mentoring, teacher release for planning and the provision of spaces for teachers to work together. Changes were also needed in matters such as space utilisation, timetabling, subject availability and budgeting. There was little

⁵⁶ It is not possible to estimate a response rate from the data available. In the trial critical friends were appointed to serve 38 schools and operated from 20 sites. .

evidence of links to the wider community. Where these links were evident it appeared to have been built on pre-existing structures where staff perceived the community to possess relevant expertise. Although critical friends were a part of all schools, the way they operated, and their impact, varied considerably. It is possible that there could be a role for critical friends in establishing the links with the wider community that were envisaged to be a feature of New Basics.

UNDERSTANDINGS, RESPONSES AND OTHER ISSUES

There was a set of research reports concerned with understandings of, and responses to, New Basics. One of these studies focused on satisfaction with trial and comparison schools using data that were gathered routinely by Education Queensland. It addressed the question of whether the changes in satisfaction over time differed between types of school. A second study examined teachers' understandings and knowledge of New Basics. A third study analysed the contributions to three discussion lists to identify the themes that were the focus and how those themes evolved over the course of the trial. The fourth used information from forums and surveys to assess wider understandings of New Basics.

School Satisfaction

Part of the research program made use of existing data to explore whether satisfaction with aspects of schooling changed for students, parents, and staff at New Basics schools.

Instrument

The analysis was based on School Opinion Surveys administered by Education Queensland each year. There were four versions of the questionnaire for primary students, secondary students, parents, staff respectively. The student and parent questionnaires consisted of six dimensions: quality curriculum, effective teaching, improved learning, public confidence, technology adoption and learning environment. Although the dimensions had the same names the items differed between primary students, secondary student, and parents. The staff questionnaire had seven dimensions: work environment, relationships, school operations, staff morale, support, work roles and work value.

Method

The regular survey was administered in all government schools with students being sampled in relation to school size from Years 5, 7, 9, and 11. Forty families were sampled for the parent or caregiver questionnaire (or all families in schools with fewer than 40 families). The staff survey was administered to all staff in 2002 and 2003. In 2000 and 2001 large schools had the option of selecting a random sample of 40 teaching staff and 40 non-teaching staff (following a set of guidelines for selection).

Data from 2000 to 2003 were used in the analysis. Thirty-four of the 38 trial schools and 67 comparison schools were included in the analysis. Comparison schools were selected from clusters established from a cluster analysis of six variables (Index of Relative Socio-Economic Disadvantage, urban/rural, band level of principal, full-time equivalent enrolment, proportion of students of Aboriginal and/or Torres Strait Islander background, type of school). For each trial school the most similar two comparison schools were selected⁵⁷.

⁵⁷ This process of selecting two comparison schools per trial school provides a better basis for comparison than selecting just one comparison school. It partly overcomes the possibility that a

In 2003 there were 1746 primary students, 1140 secondary students, 1027 parents and 1919 staff from trial schools⁵⁸. The corresponding numbers from comparison schools were 3458, 2305, 2027 and 3558. Similar numbers were used in other years.

Analysis was by multiple analysis of variance (MANOVA) for each group, using the six or seven scales as the dependent variable with year and type (trial/comparison) as the independent variables. Because it would be expected that schools might differ initially the key statistic of interest was the year by type interaction that would reflect changes in trial schools relative to changes in comparison schools.

Results

Primary students. From the main effects it was found that primary students in trial schools had higher levels of satisfaction than students in comparison schools on four of the six aspects of schooling: improved learning outcomes, public confidence, technology adoption, and safe learning environment. For all students the improved learning outcomes and technology adoption scales had increased over time. Interaction effects showed that primary students in trial schools grew relatively more satisfied than their peers in comparison schools in 2003 in the areas of quality curriculum and improved learning outcomes⁵⁹. The differences between trial and non-trial schools on these ratings were 0.10 and 0.08 of a scale unit on a five-category response scale. The differences correspond to effect sizes of 0.15 and 0.12, which are small effects.

Secondary students. There was no difference between secondary students at trial schools and those at comparison schools on any of the measured aspects of schooling. Over the course of the year all secondary students' satisfaction declined with respect to the quality curriculum and learning environment. Students from trial secondary schools had lower levels of satisfaction with the technology adoption. However, there was no significant relative change between trial and comparison schools between 2002 and 2003.

Parents. Parents from trial schools had uniformly higher values than parents from comparison schools but there was no relative change over time. There was a decline for all parents in satisfaction on the public confidence scale and an increase on technology adoption.

Staff. Staff in trial schools had higher levels of satisfaction in all areas than staff in comparison schools from the outset. There was a decrease in staff satisfaction with work roles in trial schools relative to comparison schools. There was no increase in relationships (but a tendency for a decrease) in work value, or recognition.

Interpretation

Changes in levels of satisfaction in trial relative to comparison schools were few in number and generally small in magnitude. This means that, general levels of satisfaction among primary students, parents, and staff have been maintained. In general the trial

comparison school will never perfectly match the trial school. It has the consequence that the precision of estimates for comparison schools is greater than for trial schools but this does not cause any difficulties for the statistical analysis.

⁵⁸ The numbers of trial schools involved were 29, 12, 38 and 38 for each category and the numbers of comparison schools were 50, 25, 67 and 67.

⁵⁹ The effect for learning outcomes was of borderline significance ($p=0.03$) and may have been non-significant with allowance for design effects in the sample.

neither enhanced nor reduced the satisfaction of these members of school communities. However, there was evidence that the New Basics had an effect on primary students and that they became relatively more satisfied with the quality of their curriculum and with their learning outcomes (although this was of borderline statistical significance). Teachers in trial schools were more satisfied than those in comparison schools at the beginning of the trial but the difference decreased over time.

Knowledge of New Basics

Although there was a range of documentation about New Basics, individuals constructed their own meanings.

Sources

Key messages about New Basics were identified in major documents and materials disseminated to schools. Interviews were held with members of the school community, and documents and observations were processed, to examine the knowledge the school community constructed. Interviews focused on the collective knowledge about New Basics at the school; the preparation of students for the Rich Tasks; the non-negotiable aspects of the trial; other conceptual frameworks guiding the school; the notion of exemption from other requirements; and whether the framework is seen as a trial or reality in the non-juncture years.

Method

Eight schools (six from phase 1 and two from phase 2) were chosen to provide a balanced representation of rurality and type. The disseminated messages and the site-constructed knowledge revealed in interviews were examined for congruence.

Results

There was diversity in the knowledge about, and understanding of, New Basics across the schools. In some there were several teachers who had considerable shared knowledge but in others only a limited number of teachers knew much about the New Basics. In some schools Rich Tasks were planned to be transdisciplinary while in others the tasks were developed through separate subjects. Knowledge about the non-negotiable aspects of New Basics was much less widespread than expected. In most schools there was competition between New Basics and Key Learning Areas. Teachers were uncertain as to whether that the New Basics provided an adequate base for the development of literacy and numeracy skills, and were concerned about the knowledge base in the tasks and the preparation they provided for senior school studies. There was a lack of clarity regarding exemptions from other curriculum requirements. Teachers in supportive schools tended to regard the New Basics as a trial with an opportunity to test new approaches. In other schools there was scepticism about the trial and its purposes.

Interpretation

Although the data were based on a small number of schools and teachers there were some important inferences about implementation. Knowledge about New Basics varied across schools. The variation appeared to be influenced by the conditions in each school and the culture of the school community. Schools that supported participation in reform with

leadership that was supportive of professional learning tended to construct knowledge that was more congruent with the intentions of New Basics.

New Basics Discussion Lists

Three discussion lists were established during the New Basics trial: the framework list with open participation; the New Basics project list for principals, teachers and New Basics branch staff; and the critical friends list for critical friends and New Basics branch staff.

Source

The lists were analysed in terms of topics, changes to these topics over time and whether different subgroups indicated different understandings, posed different questions or shared different resources.

Method

The report presented a descriptive account of the contributions to the lists based on qualitative analysis around some categories of school reform literature: curriculum, pedagogy, assessment, reporting, and the nature of implementation.

Results

The topics of the discussion list changed over time from conceptual issues associated with what was intended to more practical issues of implementation. The contributions by educators used the language of official documents (eg the centrality of teaching and teachers and the importance of intellectual rigour). Participation changed over time. Initially the discussions involved key figures in the reform and members of other educational institutions. Later they involved teachers and members of the New Basics branch about requirements for the trial. There were some issues that were not discussed extensively including the organisational structure of schools and the structure of teachers' work.

Interpretation

Although the report does not detail in quantitative terms the amount of use of the discussion lists, or how many different people participated, it provides an analysis of the topics covered and how those changed over time.

Responses to New Basics

Knowing the perceptions and understandings of New Basics among participants in and observers of the trial was considered important for judgements about future developments.

Sources and Methods

Several studies were conducted to ascertain different aspects of responses to New Basics. Public consultations were held in late 2000 to gain an understanding of the community's awareness and perceptions of the New Basics in seven regions where a trial school was located. These involved small groups of members of the general community and representatives of local business and government with a facilitator. Community information forums (one hour) were held in 2001 to provide information on New Basics

in 14 locations. Some 286 people (104 community, 182 educators) attended 28 forums in response to advertisements and written invitations. A market research company used a combination of telephone surveys, face-to-face interviews and focus groups to interview a range of people about New Basics. These included interviews with 69 school principals and 193 teachers from trial and non-trial schools in 2001. Four focus groups were held with 41 parents between April and July 2002. Discussions were also held with key personnel from the Queensland Department of Education and the Arts.

Results

In the community consultation groups there was little knowledge of the New Basics in the community so discussions centred around views of education in general. The issues that emerged were: basic skills, change in the world around us, employment and higher education, community involvement in the life of a school, expert transmission of knowledge to students, individual student needs, and cultural diversity. Participants in the community information forums viewed the provision of information to the community as essential and were supportive of New Basics (a majority of participants were educators). Concern was expressed about: assessment, group work, resourcing, professional development, pre-service education, school structures, postcompulsory schooling, students with special needs, and transient students. In the market research survey the teachers and principals indicated that they had obtained information about New Basics from different sources and there were other differences between trial and non-trial schools. Participants used the term “real life” to describe core elements of New Basics. The trial was associated with “hard work”, “change” and “stress”; it was viewed positively by participants from trial schools and negatively by those in non-trial schools. The value drivers for teachers and for principals were: “getting back to the business of teaching”.

Costs and Feasibility

The cost of the trial of the New Basics was estimated as approximately \$11 million (depending on the price indexing that is applied over the years of the trial). Almost all (37 out of 38) of the schools in the Phase 1 trial decided to continue to implement the essential elements of New Basics after the trial concluded in 2003. In terms of per student costs for the system, or even just for the participating schools, it was a relatively modest addition. The question of the return on the investment is harder to judge. As a result of the trial a great deal has been learned about teaching, learning and assessment in schools for the education system and by the participating schools and teachers. The form in which what has been learned is incorporated in future provisions for education cannot be known at this stage. Hence the return on the investment in the trial can only be estimated after a little time has elapsed. Certainly the research program identified areas that appear to have been successful within the trial and developments that may have wider ramifications for the system outside the trial and even beyond New Basics itself.

Summary

On the basis of the school surveys it appeared that general levels of satisfaction among primary students, parents, and staff had been maintained during the trial. Primary school students tended to become more satisfied with the quality of their curriculum and their learning outcomes. Knowledge about New Basics varied considerably partly in response to school conditions and the culture. Outside the world of schools, knowledge about

New Basics was sparse. Discussions on the various lists covered a range of topics with a general transition over time from conceptual issues associated with intentions to practical issues of implementation.

SUMMARY AND CONCLUSIONS

The Research Program

The Research Program addressed issues that were central to New Basics: student work (including Rich Tasks), teaching, assessment, achievement and school organisation. It used a variety of sources of data and mixed methods of analysis to derive a number of conclusions about the trial of New Basics.

Data Sources

The following were included among the important sources of data that were analysed as part of the program.

- Samples of the work of students on Rich Tasks and from everyday school activities.
- Results of observations of teachers in classrooms and a survey of students regarding their classroom experience.
- Records of teacher assessments of Rich Tasks and the processes of the moderated school-based assessment system.
- Measures of student achievement on three different types of achievement tests.
- Information about school organisation and structure derived from a qualitative study in 13 schools.
- A range of surveys concerned with responses to the New Basics and specific aspects of it.
- Records routinely gathered by the school system as part of its management processes.

Analysis

The research program involved mixed methods of analysis. The mix of methods went beyond the use of both quantitative and qualitative methods in that there were different forms of analysis used to address different issues and different forms of data. These methods, and the associated data sources, have been discussed through the course of the report. Two common approaches to analysis were to examine changes over time in trial schools and to compare results from trial schools and a comparison group of non-trial schools (that were defined in various ways in different studies). In a few instances the more powerful design of comparative studies of change over time was implemented.

In field studies random assignment to trial and non-trial conditions is usually not practical or possible. In addition there are difficulties in defining comparison groups because the data on which to identify the comparison group are not always available. In some instances, where it was appropriate, individual background data from students were used as part of the analysis. This provides better statistical control for differences

between trial and comparison schools and allows for more extensive analysis. More use of individual level data would have enriched the analytic possibilities. In addition, where data are from two-stage sample designs such as students within schools, measures of precision or tests of statistical significance, based on the application of simple random sample statistics, will give misleading results (indicating significance where differences are not significance). Where this is an issue it has been noted.

One of the important aspects of the research program was to articulate a multi-faceted concept of richness as being recognisable by a range of people and to use measures of that concept in studies of Rich Tasks and other samples of student work.

Scope

The Research Program was a three-year study that included varied projects and operated over the duration of the trial. Extended research programs such as this are all too infrequent in educational research but have a great deal to offer as the basis of policy decisions and to inform wider perspectives on factors that influence educational development. A necessary part of such studies is that they attend to the full range of issues in the program. Consequently other possibilities, such as a full longitudinal study of individuals moving through schools over the period, often cannot be included.

Summary Results Concerning Impact

As an aid to summarising the results from the research program, effect sizes for those studies were computed and summarised in Table 12. Two main measures of effect size are recorded in that table: standardised mean differences and the logarithm of the odds ratio. It should be noted that measures of the size of an effect are different measures of the certainty that the effect is different from zero. Where small samples are involved a large effect size may not be significant and where large samples are involved a small effect size may be significant.

Rich Tasks and Student Work

It was found that Rich Tasks in 2003 from Year 6 students in trial schools exhibited higher levels of richness than work from high achieving Year 6 students in other schools. That difference from work of high achieving students in non-trial schools was not evident at Year 3 and Year 9. In Year 3 the tendency was towards lower levels of richness in trial schools and in Year 9 the levels of richness was the same in trial and non-trial schools.

In terms of other student work there was little evidence of high intellectual quality in work collected in 2001 and 2002 but in 2003 there were differences in the intellectual quality of work from trial schools and other schools in Years 8 and 9. At those levels the effect size is large (see Table 12) but it is not clear that the difference is statistically significant. The difference appeared to be associated with the extent to which work was connected to the outside world on which dimension the differences appeared to be statistically significant.

Table 12 Effect Size Measures for Intended Impact of New Basics

Outcome	Effect Size Measure		
	Δ	Δ'	Log (ω)
Rich tasks c.f work from high achieving non-trial students: Year 3			-0.67
Rich tasks c.f work from high achieving non-trial students: Year 6			1.05
Rich tasks c.f work from high achieving non-trial students: Year 9			0.04
Intellectual quality student work, trial c.f non-trial schools: Year 2			0.04
Intellectual quality student work, trial c.f non-trial schools: Year 5			-0.04
Intellectual quality student work, trial c.f non-trial schools: Year 8			1.00
Intellectual quality student work, trial c.f non-trial schools: Year 9			1.34
Teacher observation: change in intellectual quality: 2001 to 2003	0.61		
Teacher observation: change in connectedness: 2001 to 2003	0.03		
Teacher observation: change in classroom environment: 2001 to 2003	0.35		
Teacher observation: change in recognition of difference: 2001 to 2003	0.40		
Enacted pedagogy: intellectual quality: trial c.f non trial Year 5	0.12		
Enacted pedagogy: connectedness: trial c.f non trial Year 5	0.21		
Enacted pedagogy: classroom environment: trial c.f non trial Year 5	0.31		
Enacted pedagogy: recognition of difference: trial c.f non trial Year 5	0.15		
Enacted pedagogy: intellectual quality: trial c.f non trial Year 5	0.02		
Enacted pedagogy: connectedness: trial c.f non trial Year 5	0.06		
Enacted pedagogy: classroom environment: trial c.f non trial Year 5	0.03		
Enacted pedagogy: recognition of difference: trial c.f non trial Year 5	0.08		
ISA tests Year 3: trial c.f non-trial, reading mathematics writing	-0.13		
ISA tests Year 5: trial c.f non-trial, reading mathematics writing	0.00		
ISA tests Year 9: trial c.f non-trial, reading mathematics writing	-0.07		
Statewide Literacy Tests: trial c.f non-trial (averaged Years 3, 5, 7)		0.07	
Statewide Numeracy Tests: trial c.f non-trial (averaged Years 3, 5, 7)		0.02	
School satisfaction, primary students curriculum: trial c.f non-trial	0.15		
School satisfaction, primary students learning: trial c.f non-trial	0.12		

Effect sizes are estimated as follows:

Δ = difference in means divided by pooled standard deviation

Δ' = difference in medians divided by pooled standard deviation

Log (ω) = log of the odds ratio.

Teaching and Learning in Classrooms

Observations of classroom teaching and learning practices suggested that there was a tendency towards higher scores on intellectual quality, supportive classroom environment and recognition of difference dimensions in trial compared to non-trial schools. On these three dimensions the average effect size was 0.5 standard deviations. The small number of teachers involved limits the extent that this can be regarded with certainty but the size is moderate. Student responses to the Enacted Pedagogy Assessment indicated differences between trial and comparison schools on connectedness, classroom environment and recognition of difference among Year 5 classes.

Student Achievement

There was no difference between trial and comparison schools on the International School Assessments, although there was a tendency towards less growth from March to December in trial schools at Year 3. This was not statistically significant after allowance was made for difference in other student characteristics. More detailed analysis indicated that there was no impact of being in the trial on relationships of achievement with socioeconomic background or context, or on the difference between Indigenous and non-Indigenous students. In other words on that outcome the trial had no effect on those aspects of equity but there was an effect for Indigenous students reading in Year 6. The analyses also showed greater growth over the course of Year 3 than Year 6, which was greater than in Year 9. There was a very small change in the difference between trial and non-trial schools on the statewide literacy and numeracy tests that indicated a small reduction in the gap between trial schools and all other schools.

School Satisfaction

On the measures of school satisfaction in 2003 there were higher scores on satisfaction with quality curriculum and learning outcomes for primary students in trial schools compared to non-trial schools. This difference had not been evident in 2001 and 2002 and was not evident for secondary school students.

It appeared that collective activity, supportive leadership, and openness to innovation increased during the trial and that trial schools had generally higher levels of teacher professional community than comparison schools. However, that difference may have predated the trial given that organisational capacity was a criterion for participation in the trial. It was found that scores on professional development declined in other schools but was maintained in trial schools.

Strengths and Weaknesses

Strengths

Quality of Student Work

A major strength of the New Basics program was on the nature and quality of student work. It has been noted that Rich Tasks from Year 6 students in trial schools exhibited higher levels of richness than the work of high achieving Year 6 students in other schools but that a similar difference was not evident at Year 3 or Year 9. Providing the opportunity, incentive and structure for student work of high quality represents an important strength of the trial. In Year 6, and probably more generally in the upper primary years, there seemed greater opportunity to implement and develop Rich Tasks than at the other two juncture years. It is possible that sufficient foundations in reading, number and writing needed to be established to work in the ways envisaged in Rich Tasks and that this was not feasible before Year 3. The difference for Year 9 may have been related to the difficulty of the three-year span extending into primary school and the lack of congruence between the New Basics organisers and the knowledge base of secondary school teachers and associated forms of organisation.

Development of an Assessment System

A second strength of the New Basics was the establishment of an assessment system during the trial. Although this was not part of the original conception of New Basics it

was central to establishing standards for student work and the implementation of the Rich Tasks. It is notable that evidence of richness became most evident only when the moderated assessment process was imminent. The introduction of this system provided a vehicle for reporting on the quality of student work on a wider basis than the assessment of literacy and numeracy. It may be relevant that where there were adjustments to teacher assessments as a result of moderation those adjustments were in a downward direction. That suggests that moderation was conveying some important information about standards. In Queensland, as in much of Australia, there is little assessment below the senior years that provides teachers with information about standards that they might expect from their students. It was also concluded from studying the process of the moderation system that, although teachers in the trial schools were able to assign grades to Rich Tasks competently, there was a need to develop a wider base of expertise. Queensland teachers have a stronger orientation to assessment as a means of improving learning by students. Perhaps an equally strong orientation to using assessment to guide curriculum planning and decisions about programs would be helpful in the implementation of activities such as Rich Tasks. The moderated assessment system appeared to be crucial to the implementation of the Rich Tasks in the way that they were intended. Assessment systems are inextricably linked to the implementation of changes of this sort.

Changes in Approaches to Teaching

The evidence regarding changes in teaching and learning was more mixed. The changes became evident in the third year of the trial but it is harder to establish that the changes were greater in trial than non-trial schools. The change in the third year may have been simply a matter of the time needed for the changes to take root or it is possible that this corresponded with the assessment system that made expectations clearer in more concrete terms than previously. Although there is evidence of changes over time there is no evidence linking the increased use of these strategies to student learning. Moreover, it is not clear whether the dimensions of productive pedagogies are intended to encompass a full range of teaching strategies or whether they represent aspects to be given greater emphasis than in the past. It would seem better to recognise that there are other aspects of teaching to be included and developed. Nevertheless, the New Basics emphasised wider use of certain teaching strategies and there is some evidence of that having taken place.

Weaknesses

Achievement

Measures of achievement that go beyond samples of student work are important to determine the extent to which learning can be applied in new situations and to new problems. There was no impact of being in the New Basics trial on student achievement on any of the three assessment measures. New Basics literature did not argue that there would be an impact on literacy and numeracy skills but it did argue that it would promote higher order thinking skills. The International Schools Assessment (ISA) sought to measure higher order thinking in the context of reading literacy, mathematical literacy and writing (and the analysis allowed for the effect of basic literacy and numeracy skills) and did not show a benefit from participation in the trial. It can be argued that the types of skill embodied in these tests were not those envisaged by New Basics. If that were a reasonable interpretation it would appear that there remains a need to articulate better the higher-order thinking that is envisaged by New Basics and to establish indicators that go beyond the work that students do. These indicators need not be used as part of routine assessment but as part of research studies to examine the impact of New Basics in terms

of skills and knowledge that can be transferred to and applied in new situations. The World Class Tests were intended to provide such data but for a variety of reasons did not yield information that could be used to address these issues. In terms of the intentions of New Basics it can be concluded that there was no evidence of any deleterious impact on basic literacy and numeracy as measured by the statewide tests.

Congruence with Other Aspects of the School System and its Context

The 13-school study identified the changes that schools needed to make to organisation and structure to implement the New Basics. This was especially true for large schools. Some of these changes involved the roles of teaching staff, teacher release for planning and the provision of spaces for teachers to work together. Changes were also needed in matters such as space utilisation, timetabling, subject availability and budgeting. Arguably, all of these changes are within the power of schools to change. However, there are other aspects of the context of schooling that are harder to change. One of these is the movement of students between schools. It was noted that one reason for students not being able to complete Rich Tasks was that they had transferred into a school. Rates of student movement are sufficiently high for this issue to need attention. In one sense this is an issue wherever schools and teachers have authority for curriculum but it appears to be exacerbated when the focus of student work leading to culmination as a Rich Task extends over three years. A second aspect concerned the transition to high school. The Year 9 Rich Tasks extend back to Year 7 and thus span the transition from primary to secondary school. With many high schools drawing students from multiple primary schools this created difficulties for implementing Rich Tasks in the manner intended.

A third aspect concerned the curriculum organisers of New Basics in relation to high school teaching and expectations. The curriculum organisers, intended to reflect the contemporary context in, and for which, children learn, were: life pathways and social futures; multi-literacies and communication; active citizenship; and environments and technologies. Among many teachers in schools they were seen as a desirable alternative to Key Learning Areas. These organisers were argued to be transdisciplinary and drew upon traditional disciplines (which are not the same as Key Learning Areas). Since these organisers were so much part of the fabric of what happened in schools in planning Rich Tasks it was difficult to identify through research the effects of these separately from the other attributes of trial schools. However, there did seem to be a need to explore how the New Basics and traditional disciplines could be reconciled theoretically and organisationally (Hogan, 2001). The research program indicated that the implementation of New Basics in high schools was more difficult than in primary schools. This was probably not simply due to organisational structures but reflected questions of articulation with studies in later years and articulation with the epistemological beliefs of high school teachers. There is also an issue of how the work done in and through schools can be connected to the worlds of scholarship as well as to the immediate environment of students. It is possible that much of what is done in the Year 9 Rich Tasks could reference disciplines rather than the New Basics curriculum organisers.

Differential Impact

From several perspectives there was evidence that the effects of New Basics differed between year levels. Although this is not necessarily a weakness, it is important to the evaluation of the trial and for formulating policies for the future. The impact of New Basics appeared to be greatest in the upper primary years represented by measures at Year 5 and Year 6. The specific manifestations were in richness of the Rich Tasks, in teaching and learning practices and in student satisfaction with curriculum and learning

outcomes. There was also a relatively high level of completion of Rich Tasks in Year 6. In the lower secondary school up to Year 9 there was evidence of higher quality student everyday work (especially in terms of its connectedness to the wider world) but less clear evidence of richness in relation to high achieving students from non-trial schools. In Year 9 there was a relatively high level of non-completion of Rich Tasks. There were some changes in teaching practice but to a smaller extent than was the case in the upper primary school. In lower primary school there was less evidence in the richness of student work or of intellectual quality than in the case of the upper primary or lower secondary years. There was also evidence in the unadjusted Year 3 scores (i.e. those that did not make allowance for the influence of other student characteristics) a tendency towards less growth in reading or mathematical literacy than in non-trial schools.

Equity

There is not much evidence from the trial about equity aspects of New Basics. The one aspect that could be related to individual student characteristics was the International Schools Assessment. In that there was no evidence of a different relationship between achievement and socioeconomic status in trial and non-trial schools and no evidence of differential growth related to socioeconomic status. In general there was no difference between trial and non-trial schools in the relationship between Indigenous and non-Indigenous status except for a higher level of growth in reading for Indigenous (compared to non-Indigenous) students in Year 6. The overall conclusion is that more needs to be known about the relationship of student characteristics and a range of outcomes.

Viability and Scope for Implementation

One indication of the viability of New Basics is that almost all of the trial schools agreed to continue to operate New Basics beyond the period of the trial. Of course the schools that participated in the trial had sought to be involved so their commitment to continue does not necessarily indicate that New Basics would be adopted with similar alacrity in schools that had not sought to participate. There also remain questions of benefits for students, as well as the scale, form and focus of implementation. And there are questions of how effectively a trial such as this could be extended on a wider scale. Questions of viability and scope can also involve issues of whether more general principles from the trial could be applied on a wider basis as well as questions of implementing New Basics.

One aspect of the trial that appears to be important for the operation of New Basics is that a moderated assessment system linked to the Rich Tasks is important and one of the larger ongoing costs of New Basics. The moderated assessment system appeared to be a crucial element of support to the implementation of the Rich Tasks in the way that they were intended. Assessment systems are inextricably linked to the implementation of changes of this sort. That is an important consideration for the future that should be a factor in determining how far New Basics, or aspects of New Basics might be extended. Professional development would also be important but there is only a little evidence in this report concerning professional development.

Overall the trial suggested that New Basics operated most effectively in the upper primary school and least effectively in the junior primary school. The most fruitful area for application of New Basics and New Basics principles appeared to be the upper primary school. There was mixed evidence of impacts in high schools that suggested that principles learned from the trial could find application in Years 8 and 9. However,

application in that area would probably require a review of the span of three years and possibly stronger links to the major disciplines of the curriculum. Having tasks over a period that spanned one year of primary school and two years of high school seems to have established a significant difficulty, especially for high schools whose students are drawn from a large number of primary schools. It does seem that the introduction of some of the principles of extended work tasks in high schools needs to recognise the structures of those schools, the links to the senior secondary years and the knowledge beliefs and backgrounds of the teachers.

It is not clear from the trial whether the full span of three years and mapping curriculum is necessary to achieving all the goals associated with New Basics. A framework that focussed on tasks extending over a full year might fall short of the ideals of New Basics but might be more readily implemented within schools where there is a high degree of mobility of students and teachers.

Areas for Review, Research and Development

The New Basics trial provided the opportunity to examine a wide range of issues. The research program associated with the trial investigated a great many of those issues. One of the consequences of any program of research is that it identifies further issues for investigation and highlights aspects of the program that were not previously identified as needing exploration.

There is a case to review the evidence that has been accumulated and to use that evidence to guide the adaptation of Rich Tasks to environments where implementation proved difficult. There is evidence of high schools having adapted Rich Tasks to fit the constraints that they saw as operating. In many respects the Rich Tasks are the primary focus of the New Basics but it could be that there are elements in those tasks that could find application in other environments. In any event there will be an ongoing need to develop and refine the Rich Tasks.

In the trial there was some evidence of changes in teaching and learning strategies associated with the trial that became evident in its third year. A question that arose was whether this was a result of the rich task assessment process in that year. In addition data from the trial were not as conclusive as would be wished because of limitations of sample size. It would be useful to replicate the teaching strategies study on a wider basis so that data on teaching could be linked to student outcome measures that captured change over given period of time. In such a study it would be valuable to code some other dimensions of classroom teaching, perhaps as lower inference measures, that are defined more independently of the New Basics.

New Basics literature discusses the Rich Tasks as being designed to engage students. It would be useful to investigate the extent to which students become engaged in those tasks and how they become engaged (Frederick et al, 2004)⁶⁰. Such a study could examine initial reactions to the tasks (how and why students became engaged in the task)

⁶⁰ Frederick et al (2004) characterise different forms of engagement: behavioural, emotional and cognitive. Behavioural engagement refers to participation in academic, social and co-curricular activities. Emotional engagement refers relations with teachers, fellow students and the school as an institution. Cognitive engagement refers to involvement and effort in work in trying to comprehend complex ideas and difficult skills.

and how interest was maintained during the course of the exercise. It seems likely that tasks would vary in their interest for different groups of students.

Analysis of the Rich Task assessments conducted with data from the moderation process indicated that there was a common element that accounted for about half the variation in grades in addition to features unique to each Rich Task. It is important to understand better the nature of that common element and of the clusters among the unique elements identified in the research report. This task is related to the issue of articulating better the nature of the higher order thinking skills and mapping what is meant by growth in those skills. Higher order thinking skills are sometimes written about as if they constituted a single dimension. That seems an unlikely proposition but it is important to understand the multidimensional character of higher order thinking.

In Conclusion

The New Basics research program has produced important findings from its study of the trial. Those findings go beyond a simple question of whether the New Basics worked. Implementing a multifaceted innovation such as New Basics usually means that some features have larger impacts than others and that the impacts on some outcomes will be greater than on others. The results of a set of mutually supporting studies did indicate that a concept of richness could be defined and identified and there was an impact on the richness of student work in Year 6. Those results suggest that there are benefits from having extended tasks in the style of Rich Tasks in the upper primary years. Other studies from the research program suggest that a moderated school-based assessment system is important to ensure that the tasks are implemented as intended and that common standards underpin the assessments. Given that such systems are costly it may well be appropriate to focus on tasks in Year 6 or Year 7 supported by an assessment system at that level. There was a more varied impact in high schools possibly due to a mismatch with organizational structures, curriculum and more fundamental issues. Attention could be given to modifying extended tasks to fit better within those constraints.

REFERENCES

- Agresti, A. (2002). *Categorical data analysis*. Hoboken, New Jersey: Wiley-Interscience.
- Australian Council for Educational Research (ACER) (2002). *Information for the International Schools Assessment*. Melbourne: ACER.
- Berends, M., Bodilly, S. J., & Kirby, S. N. (2002). The future of whole-school designs: Conclusions, observations, and policy implications. In M. Berends, S. J. Bodilly & S. N. Kirby (Eds.), *Facing the challenges of whole-school reform: New American Schools after a decade* (pp. 143–153). Santa Monica, CA: RAND.
- Brown, G. T. (2002). *Teachers' conceptions of assessment*. Unpublished doctoral thesis, University of Auckland, Auckland, NZ 2002
- Bryk, A. S., Nagoada, J. K., & Newmann, F. M. (2000). *Chicago classroom demands for authentic intellectual work: Trends from 1997-1999*. Retrieved 20 May, 2003, from <http://www.consortium-chicago.org>
- Burkhardt, H. (2002). World class assessment: Principles, practice and problem solving. In C. Richardson (Ed.), *Assessing Gifted and Talented Children*. London: QCA.
- Cooper, H. & Hedges, L. (1994). *The Handbook of Research Synthesis*. New York: Russell Sage Foundation.
- Cuban, L. (1998), 'How schools change reform: Redefining reform success and failure', *Teachers College Record*, 99, 453–477.
- David, H. A. (1988). *The method of paired comparisons*. New York: Oxford University Press
- Earl, L., Watson, N., Levin, B., Leithwood, K., Fullan, M., & Torrance, N. (2003a). *Watching and learning 3: Final report of the external evaluation of England's literacy and numeracy strategies*. Toronto: Ontario Institute for Studies in Education, University of Toronto.
- Earl, L. (2003b). *Assessment as Learning: Using Classroom Assessment to Maximize Student Learning*. Thousand Oaks, CA: Corwin Press.
- Earl, L. (2004). *Assessment as learning*. Presentation to the Incorporated Association of Registered Teachers in Victoria. Melbourne: IARTV .
- Education Queensland, (1999). *Queensland State Education 2010*. Brisbane: Department of Education.
- Education Queensland (2000). *New Basics Technical Paper*. Brisbane: Department of Education.
- Education Queensland (2001a). *New Basics: Theory into Practice*. Brisbane: Department of Education.
- Education Queensland (2001b). *New Basics: Curriculum Organisers*, Brisbane: Department of Education.
- Education Queensland (2001c). *New Basics: The why, what, how and when of Rich Tasks*. Brisbane: Department of Education.
- Education Queensland. (2003a). *New Basics—Research program*. Brisbane: Department of Education.
- Education Queensland. (2003b). *The Department of Education strategic plan 2003–2007*. Brisbane: Department of Education.
- Fleiss, J. (1994). Measures of effect size for categorical data. In H. Cooper and L. Hedges, L. (1994). *The Handbook of Research Synthesis*. New York: Russell Sage Foundation, pp 245-260.
- Feldt, L. S., & Brennan, R. L. (1993). Reliability. In R. L. Linn (Ed.), *Educational Measurement* (pp. 105–146). New York: The Oryx Press.

- Frechtling, J., & Sharp, L. (1997). *User-Friendly Handbook for Mixed Method Evaluations*. Washington DC: National Science Foundation, Directorate for Education and Human Resources.
- Frederick, J., Blumenfeld, P., Paris, A. (2004). School engagement: potential of the concept, state of the evidence. *Review of Educational Research*, 74 (1), 59-120.
- Fullan, M. (1993). *The New Meaning of Educational Change*. New York: Falmer Press.
- Fullan, M. (1995). The school as a learning organization: Distant dreams. *Theory into practice*, 34(4), 230–235.
- Glass, G., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in Social Research*. Beverly Hills, CA: Sage.
- Hogan, D. (2001). *Review of NELS*. Hobart: Faculty of Education, University of Tasmania
- Hummel-Rossi, B., & Ashdown, J. (2002). The state of cost-benefit effectiveness analyses in education. *Review of educational research*, 72(1), 1–30.
- Kellaghan, T., Stufflebeam, D. & Wingate, L. (2003). Introduction. In T. Kellaghan and D. Stufflebeam (Eds.) *International Handbook of Educational Evaluation*. Dordrecht, The Netherlands: Kluwer.
- Light, R. & Pillemer, D. (1984). *Summing Up: The Science of Reviewing Research*. Cambridge, Mass: Harvard University Press.
- Louis, K. S., Marks, H. M., & Kruse, S. D. (1996). Teachers' professional community in restructuring schools. *American educational research journal*, 33(4), 757–798.
- Matters, G. (2002). The Rich Task: A new assessment species for new times. In C. Richardson (Ed.), *Assessing gifted and talented children* (pp. 35–49). London: Qualifications and Curriculum Authority.
- Matters, G. (2004). Growing an assessment culture. *Assessment and reporting*, 3, Brisbane: Department of Education.
- Newmann, F. M., & Wehlage, G. G. (1993). Five standards of authentic instruction. *Educational Leadership*, 50(7), 8–12.
- Newmann, F. M., Marks, H. M., & Gamoran, A. (1995). *Authentic pedagogy: Standards that boost student performance* (Report No. 8, Issues in restructuring schools). Madison, WI: University of Wisconsin, Center on Organization and Restructuring of Schools.
- Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.
- Organisation for Economic Cooperation and Development (OECD) (1999). *Measuring Student Knowledge and Skills: A New Framework for Assessment*. Paris: Organisation for Economic Cooperation and Development.
- Queensland Studies Authority. (2003). *Queensland Years 3, 5 and 7 tests in aspects of literacy and numeracy: 2003 test reporting handbook*. Brisbane: Queensland Studies Authority.
- Ross, K. & Wilson, M. (1994). Sampling errors in survey research. In T. Husen and T. N. Postlethwaite (Eds). *International encyclopaedia of educational research*. (Vol 9, pp 5124-5131). London: Pergamon.
- Ross, K. & Rust, K. (1994). Sampling in survey research. In T. Husen and T. N. Postlethwaite (Eds). *International encyclopaedia of educational research*. (Vol 9, pp 5131-5142). London: Pergamon.
- Rowe, K. J. & Hill, P. W. (1996). Assessing, recording and reporting students' educational progress: The case for 'subject profiles'. *Assessment in education*, 3(3), 309–352
- Scriven, M. (1967). The methodology of evaluation. In R. E. Stake (Ed). *Curriculum Evaluation*. AERA Monograph Series on Evaluation, Chicago: Rand McNally.

-
- Stoel, C. & Anderson, E. (2000). Broadening Teaching Perspectives, *Basic education*, 44, (9), 1-7. <http://www.c-b-e.org/be/iss0005/a0toc.htm>.
- Tashakkori, A. and Teddlie, C. (Eds.) (2002). *Handbook of mixed methods in social and behavioural research*. Thousand Oaks, CA: Sage Publications.
- The University of Queensland, School of Education. (2001a). *The Queensland school reform longitudinal study*. Brisbane: The State of Queensland (Department of Education).
- The University of Queensland, School of Education. (2001b). *The Queensland school reform longitudinal study: Supplementary material*. Brisbane: The State of Queensland (Department of Education).