Teachers' Conceptions of Assessment

Gavin Thomas Lumsden Brown

M. Ed. (Hons.) Massey University

B. Ed. (TESL) (Dist'n) Concordia University

Supervisor: Professor John A. C. Hattie, University of Auckland

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Education, The University of Auckland, 2002

Abstract

Teachers' conceptions are powerful in shaping the quality of their instructional practice. The purpose of this thesis is to defend a four-facet model of teachers' conceptions of assessment, which revolves around emphasising improvement or school accountability, or student accountability purposes or treating assessment as irrelevant. Further, it explores how those conceptions relate to teachers' conceptions of learning, teaching, curriculum, and teacher efficacy.

A literature review is used to identify the major conceptions. Multiple studies led to a 50-item Teachers' Conceptions of Assessment (COA-III) questionnaire based on the four main conceptions of assessment. Structural equation modelling showed a close fit of a hierarchical, multi-dimensional model to the data. Teachers moderately agreed with the improvement conceptions and the system accountability conception. Teachers disagreed that assessment was irrelevant. However, teachers had little agreement that assessment was for student accountability. Improvement, school, and student accountability conceptions were positively correlated. The irrelevance conception was inversely related to the improvement conception and not related to the system accountability conception.

A four-factor structure of teachers' beliefs about assessment, curriculum, teaching, learning, and teacher efficacy, was found. Teachers agreed that assessment influences and improves their teaching and student learning. They agreed less strongly that assessment, measuring surface learning only, makes schools, teachers, and students accountable and that teachers are able to conduct assessment through a systematic technological approach. They agreed at a similar level with student centred learning that involves deep approaches to learning, divorced from assessment. They disagreed with a telling type of teaching that focuses only on intellectual development of students or on reconstruction or reform of society.

Use of the CoA-III makes teachers' conceptions of assessment more explicit and will assist in the development of teacher training programs, the design of assessment policy, and enhance further research into educational assessment practices. Furthermore, explicit attention to teachers' conceptions of assessment is expected to be a precursor to teachers' self-regulation of their assessment beliefs and practices.

Acknowledgements

Table of Contents

Tables

Figures

Introduction

Understanding and communicating assessment information has been identified as one of the most critical teacher professional development needs in the field of educational measurement expertise (Plake, Impara, & Fager, 1993). Furthermore, understanding what teachers believe about assessment is critical in designing and implementing appropriate teacher professional development. Tittle (1994, p. 161) has proposed that understanding teachers' conceptions of assessment is crucial in implementing assessment systems; "test theory and validity theory need expansion to encompass … theories of teaching, learning, curriculum, and change, and to view the interpreter and user as central to the development and evaluation of assessments". Ekeblad and Bond (1994) argued, in the process of advocating for phenomenographic research, that changes to the quality of student learning outcomes may depend on research that shows the internal relations between persons and things experienced; in this case the relationship of teachers to assessment.

Teachers' beliefs are understood as being organised into systems wherein some beliefs are more central or primary and others may be derived or peripherally linked to those central beliefs (Pajares, 1992; Thompson, 1992). Beliefs are the meanings connected to psychological objects or phenomena that are used as a culturally defined lens through which sense is made of events, people, and interactions (Pratt, 1992a) and which are contingent upon environmental contexts that define the objects and sense makers (Ekeblad & Bond, 1994). A wide variety of language has been used to refer to teachers' beliefs, including 'teachers' subjectively reasonable beliefs' (Harootunian & Yarger, 1981), 'untested assumptions' (Calderhead, 1996), and 'implicit theories' (Clark & Peterson, 1986).

However, in order to more efficiently address the differences between and relationships among knowledge, beliefs, belief systems, and belief clusters and the varying terminology referring to beliefs, Thompson (1992, p. 130) invoked conceptions "as a more general mental structure, encompassing beliefs, meanings, concepts, propositions, rules, mental images, preferences, and the like". Furthermore, conceptions represent different categories of ideas held by teachers behind their descriptions of how educational things are experienced (Pratt, 1992a). Thus, conceptions act as a framework though which a teacher views, interprets, and interacts, with the teaching environment (Marton, 1981). It is in this manner that a wide view of conceptions as an organising framework by which an individual understands, responds to, and interacts with a phenomenon is taken in this research. Nevertheless, the structure of teachers' conceptions is not uniform and simple; conceptions appear to be multi-faceted and interconnected. Indeed, findings exist that suggest that, "the conceptions that teachers do hold … do not fit neatly into the research-based typology and that they may be more complex and eclectic than those of … researchers" (Clark & Peterson, 1986, p. 287) and that there are "wide variations in teachers' belief systems" even when teachers share a commitment to similar educational practices (Clark & Peterson, 1986, p. 289).

The study of teachers' conceptions of assessment is important because evidence exists that teachers' conceptions of teaching, learning, and curricula influence strongly how teachers teach and what students learn or achieve (Calderhead, 1996; Clark & Peterson, 1986; Pajares, 1992; Thompson, 1992). Indeed, teachers' beliefs about student self-confidence, morale, creativity, and work are, as Asch (1976, p. 18) argued, "closely linked to one's choice of evaluation techniques". Tittle (1994, p. 151) proposed that teachers "construct schemas or integrate representations from assessments

into existing views of the self, of teaching and learning, and of the curriculum, broadly construed". From their survey of elementary school teachers, Cizek, Fitzgerald, Shawn, and Rachor (1995) argued that, based on the highly individualistic nature of assessment practices, many teachers seem to have assessment policies based on their idiosyncratic values and conceptions of teaching. In a study of high school English classes, Kahn (2000) has argued that teachers used a wide variety of seemingly conflicting assessment types because they eclectically held and practiced transmission-oriented and constructivist models of teaching and learning.

Thus, all pedagogical acts, including teachers' perceptions of and evaluations of student behaviour and performance (i.e., assessment), are affected by the conceptions teachers have about their own confidence to teach, the act of teaching, the nature of curriculum or subjects, the process and purpose of assessment, and the nature of learning among many educational beliefs. It is critical that the conceptions teachers have about teaching, learning, assessment, curriculum, and teacher efficacy and the relationships of those conceptions among and between each other are made explicit and visible. This is especially so, if it is considered prudent or advisable that teachers' conceptions be changed (Borko, Mayfield, Marion, Flexer, & Cumbo, 1997).

The New Zealand Context

Environmental constraints and opportunities play an important role in any model of teachers' thoughts and actions (Clark and Peterson, 1986; Dahlin, Watkins, & Ekholm, 2001). It is important to briefly describe the New Zealand educational context because it may be significant in understanding teachers' conceptions. In the last decade and a half large structural changes have been initiated in education. Prior to these changes, described below, New Zealand had a centralised, national, bureaucratic system

wherein all things educational (e.g., policy, curriculum, school infrastructure, teacher training, teachers' salaries, school and teacher inspection, student leaving qualifications, and school governance) were determined for primary schools by the Department of Education (Crooks, 2002).

In the late 1980s, a significant change to the administration of education was implemented under the title *Tomorrow's Schools*. The present New Zealand Ministry of Education (MoE) was created out of the old Department of Education as a policy only body. Other statutory bodies were created to deal with important functions devolved from the MoE; specifically, the Education Review Office (ERO) was made responsible for quality assurance of schools, the New Zealand Qualifications Authority (NZQA) was made responsible for secondary and tertiary level qualifications, and, perhaps most radically, all schools were made responsible for their own administration and management, through single-school boards (Wylie, 1997).

In the early 1990s, the MoE created the New Zealand Curriculum Framework (NCF) of seven essential learning areas (i.e., Language and Languages, Mathematics, Science, Social Science, Physical Well-being and Health, Technology, and Arts) each of which was broken into eight hierarchical levels of achievement covering Years 1-13. At the same time the New Zealand Qualifications Framework (NQF) was instituted which, like the NCF, has eight levels of certification ranging from Level 1 (School Certificate at the time of this research is being replaced by the National Certificate of Educational Achievement 1) expected in Year 11 (approximately age 16) of schooling to Level 8 (awarded as an equivalent to a doctorate degree). The goal in the NQF and NCF policy developments was a seamless education system that wove together curriculum and qualifications from childhood to adulthood. In addition, legislation (the National Educational Goals and National Administrative Guidelines) was enacted that

4

required schools to ensure that students reached expected levels of achievement, especially in literacy and numeracy. It was a particular responsibility of the ERO to verify that schools were complying with this legislation. Fiske and Ladd (2000) provide an excellent review of the impact of this restructuring on New Zealand schools.

New Zealand schools are generally either in the Primary (Years 1-8) or Secondary (Years 9-13) sectors each having different employment arrangements through collective employment contracts with two different teacher unions. The Primary sector consists of three major school types: Contributing (Years 1-6), Intermediate (Years 7-8), and Full (Years 1-8). Though it should be noted that some secondary schools have attached or integrated intermediate school departments. The greatest proportion of students progress through contributing and intermediate schools before commencing secondary schooling (Crooks, 2002). Students begin Year 1 at age five and the nominal age of students in each year is determined by adding that value to each school year. The MoE determines the socio-economic status of schools by considering income and ethnicity characteristics of each school community. The socio economic status of the cohort of students for each school is identified by its decile with ten being the highest and one the lowest.

Despite the external accountability role of the ERO, there is no centrally mandated testing of students in primary schools (Crooks, 2002). Assessment in primary schools is voluntary and Crooks (2002) provides a good description of the various assessment issues and practices around primary school use of teacher-made observations, conversations, checklists, and standardised tests. In the context of self-managing schools, assessment practices are school-based. High proportions of schools have reported use in at least one class of the voluntary, standardised *Progressive Achievement Tests* of language skills while only half reported using the same series'

5

mathematics tests (Croft & Reid, 1991). More recently, it was found that a large number of standardised achievement and diagnostic assessment tools were being used in New Zealand primary schools with most teachers reporting that the use of voluntary diagnostic assessments frequently or always altered the way they taught their students (Croft, Strafford, & Mapa, 2000). The stake or consequence of these school-based assessments is low with teachers assessing their students to identify their strengths and weaknesses in progress towards curriculum objectives and to evaluate the quality of teaching programmes (M. Hill, 2000a).

Assessment in secondary schools is also voluntary, but nearly every student participates in one or more of the various national qualifications assessments in Years 11, 12, and 13. These high-stakes qualifications, determined by the National Qualifications Framework, include the New Zealand Qualification Authority's external 3-hour subject examinations at School Certificate (Year 11) and School Bursary (Year 13) levels, internally assessed unit standards towards National and Sixth Form Certificates (Year 12), and the new National Certificate of Educational Achievement standards (Crooks, 2002 provides further details of the system).

Several qualitative, case-study descriptions of teachers' understandings of assessment in the context of professional development in assessment have been reported in New Zealand (Cowie, 2000; Dixon, 1999; Hill, J., 1998; Hill, M., 2000a) and will be referred to in the review of literature.

Structure of Thesis

This thesis grew out of an interest in how teachers teach and how students learn in the context of preparation for external high-stakes and internal low-stakes examinations. The impact of the assessment regime in secondary schools (documented

6

in part in Study 1) led to a review of the literature on what teachers believe about assessment, beyond what textbooks on educational measurement and classroom assessment argued teachers ought to believe about assessment. That review found little empirical information on how teachers conceive of assessment, and thus began research into the development of a tool to obtain insight into teachers' conceptions of assessment. It was expected that a variety of conceptions would be found.

Thus, the purpose of this research was to investigate how teachers conceive of assessment, and then to ascertain the structure of those conceptions, and how those various conceptions of assessment related to each other. Also, it was intended to determine the strength of agreement that teachers might have for each conception by obtaining a sample representative and large enough to make generalisations about teachers' conceptions of assessment. Further, it was intended to gain an understanding of how teachers' conceptions of assessment relate to their conceptions of learning, teaching, curriculum, and teacher efficacy. The common methodological approach taken in all the studies in this research is that of attitude scale development which involves both qualitative and quantitative approaches. Specifically, this involved generating statements based on literature and research related to the topic of interest, getting participants to indicate degree of agreement, factor analysis of responses to detect mathematical patterns that could be meaningfully interpreted in light of the literature, and verification of the factor structure through simultaneous equation modelling.

Chapter one reviews the literature on conceptions of learning and assessment and briefly overviews conceptions of curriculum, teaching, and teacher efficacy. Models of learning, curriculum, teaching, and teacher efficacy were adopted from

previous research, while a four-facet model of teachers' conceptions of assessment was developed.

The second chapter presents two studies in which primary and secondary school teachers' conceptions of learning, teaching, and assessment were explored. The approach used to obtain data from teacher self-reporting of beliefs (e.g., response scale format, avoidance of negatively worded items) was developed in these studies. The first study clearly pointed to the important role of assessment in the life of teachers and the second identified a range of beliefs about assessment.

The studies reported in chapter two provided a basis for further studies wherein an instrument to measure teachers' conceptions of assessment was researched and developed. Thus, chapter three takes up the development of a model of and an instrument to measure teachers' conceptions of assessment in a series of three studies. A Likert-style self-report instrument was administered to over 500 primary school teachers and a measurement model analysis of the data led to a robust hierarchical, intercorrelated model of teachers' conceptions of assessment. That tool was then used to shed light on how teachers' beliefs about learning, curriculum, teaching itself, and their efficacy as teachers related to their conceptions of assessment.

The fourth chapter analyses the relationship of teachers' conceptions of assessment with their conceptions of curriculum, learning, teaching, and teacher efficacy and with their self-reported assessment practices. By treating the various scale scores as observed variables a further meta-level measurement model was developed. That model reduced teachers' conceptions to four main factors (i.e., external checking, assessment influences learning, student centred learning, and telling for change).

The last chapter summarises methods and findings and discusses possible implications of the data. From a series of inter-connected studies, has grown an

integrated understanding of what assessment means to New Zealand teachers and which

acts as a basis for future research into the implication of those beliefs for teaching and

assessment practice. The chapter concludes with possible directions for future research

based on the findings and hypotheses generated by this research.

CHAPTER I. TEACHERS' INSTRUCTIONAL CONCEPTIONS: LEARNING, CURRICULUM, TEACHING, EFFICACY, AND ASSESSMENT

This chapter explores five key teachers' instructional conceptions that impact on educational achievement, those of learning, curriculum, teaching, efficacy, and assessment. Calderhead (1996, p. 719) argued that there were five main areas in which teachers have been found to hold significant beliefs (i.e., beliefs about learners and learning, teaching, subjects or curriculum, learning to teach, and about the self and the nature of teaching) and noted that "such areas, however, could well be interconnected, so that beliefs about teaching, for instance, may be closely related to beliefs about learning and the subject". It is the argument of this thesis that teachers' conceptions of assessment are an important part of this potentially interconnected set of teachers' instructional beliefs that effect students' learning outcomes. These beliefs constitute a set of value premises from which decisions about curriculum objectives, content, organisation, teaching strategies, learning activities, and instructional assessment are made (Cheung & Ng, 2000). Thus, it is expected that what teachers believe about the purpose of teaching and learning, what subjects should be taught, and how teachers might be effective may be important to understanding teachers' conceptions of assessment.

This chapter reviews the literature on these instructional conceptions and gives special attention to teachers' conceptions of assessment as it was found that those conceptions were the least well understood. First, teachers' conceptions of learning are reviewed. Learning is conceived of as a cognitive process involving remembering and understanding material, which can range from surface to deep. Then, teachers' conceptions of teaching and curriculum are reviewed and it is found that those

conceptions vary along a continuum of teacher - student orientations. Next, teacher efficacy conceptions are reviewed and it is shown that these are either an internal confidence to effect student learning or an emphasis on external obstacles that prevent teachers from effecting student learning. Finally, assessment is conceived around the three purposes of improvement of learning, accountability of students, and accountability of schools, to which teachers have varying conceptions that lead acceptance or rejection.

Conceptions of Learning

A powerful model for understanding how teachers conceive of learning is the surface-deep continuum developed in the last quarter of a century by researchers in Scandinavia (Marton and Saljö, 1976), Australia (Biggs, 1987a), and Britain (Entwistle, 1997). Marton's phenomenographic work (1981) focused attention on what students or learners claimed as their intention or purpose for learning and the processes by which the learning intention was carried out (Entwistle & Marton, 1984). A taxonomy of learning views was developed that took account of the various surface and deep ways people had of understanding learning. The surface approaches or conceptions included a) remembering things, b) getting facts or details and c) applying information. In contrast, the 'deep' approach to "learning is a qualitative change in one's way of understanding some aspect of reality" (Marton, 1983, p. 291). The deep views included d) understanding new material for oneself without reference to rewards, e) perceiving or understanding things in a different and more meaningful way, and f) developing or changing as a person.

Biggs (1987a) identified in his 3P (i.e., presage, process, product) model of learning that what students bring (presage factors such as student ability, home and

family characteristics, and other socio-demographic variables) to learning impinged on the student's selection of learning processes and purposes, as well as the products created by the student. There were three major kinds of learning purposes: surface purposes involved accurate reproduction of material; deep purposes emphasised making meaningful connections; and achieving purposes had to do with maximising rewards such as scores, or rank positions. In his process strategies, Biggs and Collis (1982) identified a surface to deep taxonomy of learning products, called the Structure of Observed Learning Outcomes (SOLO). The SOLO taxonomy described student learning as being unistructural (i.e., focused on one idea or piece of information), multi-structural (i.e., focused on more than one idea or piece of information handled serially without reference to any connection between ideas or information), relational (i.e., focused on the interconnection between two or more ideas or bits of information), and extended abstract (i.e., focused on the general principles underlying ideas or information that can be extracted from data).

Thus, with both the Marton and Biggs models, the surface to deep conception can be used to describe beliefs about the nature of learning, the purposes for learning, the processes of learning, and the products of learning. The deep intention is to understand ideas for oneself and is achieved by transforming information. The surface intention emphasises coping with course requirements and is fulfilled by consuming or reproducing information. Further, the surface approach to learning involves learners applying teachable skills such as underlining, mind mapping, or mnemonics. A third approach, achieving, has been identified that makes use of surface methods for a relatively surface intention of obtaining the highest possible grades through concerted organisation of time, space, materials, and information.

The 'surface' approach to learning is associated with the act of reproducing information that has been attended to, stored in, and retrieved from memory; for example, "in situations where the learner's aim is to gain new information or add to their store of knowledge" (Howe, 1998, p. 10). MacKechnie and MacKechnie (1999) found, in a review of programs at one New Zealand university for students entering with an academically insufficient background, that the strategies focused on were largely achievement-improvement skills such as note taking, time and study management, library skills, and reading skills. Anthony (1994, 1997) has described the use of learning strategies of Sixth Form mathematics students where she noted that the surface-oriented requirements of assessment and students' resistance to engaging in self-regulated construction of knowledge resulted in a passive approach to learning. She further found that, because of an emphasis on surface approaches, "the range and quality of many students' strategic behaviours were limited, and in many cases ineffective in assisting their knowledge construction processes" (Anthony, 1997, p.2).

The difficulty with an over-emphasis on surface approaches to learning is that learning is often a '*less-structured*' task that "cannot be broken down into a fixed sequence of subtasks or steps that consistently and unfailingly lead to the desired end result" (Rosenshine, Meister, & Chapman, 1996, p.18). Further, learning "implies a deliberate effort by the student to understand, remember, and use specified knowledge or procedures" (Devine, 1991, p. 743). Additionally, Gall, Gall, Jacobsen, and Bullock (1990, p. 10) emphasised that learning involves a variety of study skills; that is, "*the effective use of appropriate techniques for completing a learning task*" [italics in original], though it needs to be noted that there is little evidence for this claim.

While surface and deep conceptions of learning and studying can be invoked, too often learners and teachers assume that surface beliefs and approaches are bad while

13

deep ones are good. However, just as surface approaches alone cannot ensure learning

takes place, deep approaches alone cannot guarantee learning. Learning at all levels

requires active mental processing of information, the making of meaningful connections

between and among ideas and information, and repetition, practice, and memorisation

(Howe, 1998). Thus, learning requires deep (i.e., active processing of information to

make meaningful connections) and surface (i.e., use of rehearsal and repetition)

practices and processes. Successful learners seem to understand that both surface and

deep processes are legitimately involved in learning and are able to select and

implement appropriate strategies (Purdie & Hattie, 1999).

A major factor that affects student learning is the nature of assessment that

generates learning outcomes or products and the resulting feedback. Learning is

affected by assessment systems, though perhaps more often negatively than otherwise

(Crooks, 1988). Success in learning depends on the type of assessment being prepared

for (Pressley, et al., 1997). Surface approaches to learning seem eminently sensible for

assessments for which one has only overnight to prepare (e.g., which of the following is

a planet between earth and the sun?), while deep approaches seem necessary for the

completion of multi-year innovative research. As a result of assessment students

usually receive feedback. External feedback from sources outside the student (e.g.,

scores, grades, or teacher comments) in response to learning outcomes contributes

somewhat to the learner's understanding of the quality of learning processes and

products. However, it has been argued that internal monitoring of the quality of one's

own learning processes and products (i.e., metacognitive evaluation) contributes even

more than external feedback to successful learning (Butler & Winne, 1995).

How assessment is conceived, and how those conceptions relate to conceptions

of teaching, curriculum, and learning is relatively unexplored (Dahlin, Watkins, &

Ekholm, 2001). Recent phenomenographic research in Sweden and Hong Kong has identified that tertiary lecturers' conceptions of assessment impact on their understandings about student motivation, curriculum content, student ability, and student learning strategies (Dahlin, Watkins, & Ekholm, 2001). That research proposed a continuum wherein assessment was understood as moving from an external relation to teaching, learning, and curriculum until there is "a completely internal relation between curriculum and assessment" wherein "understanding, reflecting, interpreting, analysing, and relating" are essential to teaching, curriculum, student learning, and assessment (Dahlin, Watkins, & Ekholm, 2001, p. 69). In a micro-study of three mathematics teachers, Delandshere and Jones (1999) argued that teachers' beliefs about assessment are shaped by how they conceptualise learning and teaching.

Conceptions of Curriculum

Studies have explored how teachers conceive of teaching various subjects, including mathematics, English, reading, language, history, and social studies (Calderhead, 1996; Clark & Peterson; Thompson, 1992). These studies have shown that teachers develop a subject understanding that is "broad and deep, enabling them to facilitate the building of similar connections in the minds of others" (Calderhead, 1996, p. 716). They also have shown that the way teachers understand their subject affects the way they teach and assess. For example, in the field of mathematics, different major conceptions of the subject (i.e., relational understanding and instrumental understanding) are claimed to be "at the root of disagreements about what constitutes 'sound' approaches to the teaching of mathematics and what constitutes 'sound' student assessment practices" (Thompson, 1992, p. 133). In particular, those who conceive of mathematics in relational terms appear to emphasise authentic, problem-solving

process-focused forms of assessment, while those who conceive of mathematics in instrumental terms seem to emphasise correct answer-focused forms of assessment.

However, primary school teachers are not subject specialists for the most part; they are generalists charged with responsibility for teaching all essential learning areas. This indicates that examining how teachers conceive of the totality of school curriculum instead of the separate subjects taught in classrooms is important. Curriculum has to do with the answers to such commonplace questions as "what can and should be taught to whom, when, and how?" (Eisner & Vallance, 1974). Teachers can be viewed as simply delivery mechanisms or conduits for curriculum or else they can be understood to be creators or makers of curriculum (Clandinin & Connelly, 1992). It is in the spirit of the latter type that research into how teachers conceive of curriculum has been conducted. Furthermore, where teachers are concerned with curriculum-based assessment, as they are in the New Zealand context of curricula defined by eight levels of achievement objectives and legislation that requires them to monitor student progress against such objectives and levels, the orientation teachers have to curriculum may impact on what they believe about and how they use assessment. For example, teachers who believe curriculum is about transmission of traditional academic knowledge may well believe assessment is about student accountability and, thus, tend to agree with the use of surface-oriented, factual-recall, high-stakes, externally referenced, objectively scored assessments.

At least five major orientations to curriculum have been found in a review of the literature on what students should be taught (Eisner & Vallance, 1974). These orientations included (a) the development of cognitive processes, (b) the finding of efficient technological means to a set of agreed ends, (c) the primacy of self-actualisation, (d) the importance of social reconstruction and relevance, and (e)

16

traditional academic rationalism (Eisner & Vallance, 1974). While these conceptions of curriculum have a different focus to those that interpret curriculum primarily in terms of the politics of defining what is taught (e.g., curriculum as a fact, as practice, or as social conflict in Goodson, 1995) or understanding the nature of what is taught (e.g., curriculum as race, gender, aesthetic, institutionalised, or poststructuralist texts in Pinar, Reynolds, Slattery, & Taubman, 1995), they remain the focus of interest in this thesis.

Cheung (2000), found in a review of orientation to curriculum models that there were common elements in the various models, among which were the five constructs identified by Eisner and Vallance (1974). Specifically, they were (a) cognitive processes or skills, (b) the role of technology, (c) society and social change, (d) humanistic concern for individual development, and (e) academic knowledge or intellectual development. Like other research into teachers' beliefs, Cheung (2000) has argued that these orientations to curriculum (a) explain why teachers emphasise certain topics, (b) clarify the real meaning or intent of curriculum documents, and (c) influence both teacher professional and curriculum development. Inspection of curriculum practice is not guaranteed to expose teachers' true orientation to curriculum as various contextual constraints may impose common curriculum practices on teachers with highly divergent views of curriculum (Cheung & Ng, 2000). Although teachers have interconnected conceptions of curriculum drawing on several orientations simultaneously, there appear to be patterns in teacher conception of curriculum (Cheung, 2000).

Cheung (2000) operationalised one model of curriculum conceptions that has four major orientations (i.e., humanistic, social reconstructionistic, technological, and academic) into a teacher self-report instrument. The humanistic conception advocates that the student is the crucial source of all curriculum, the social reconstructionist

perceives school as a vehicle for directing and assisting social reform or change, the technological orientation focuses on finding efficient means of reaching planned learning objectives through the use of modern technology, and the academic orientation aims at developing students' rational thinking and skills of inquiry. Cheung's (2000) research with Chinese speaking, Hong Kong primary school teachers found that the highly inter-correlated technological and academic orientations most strongly explained teacher conceptions of curriculum, closely followed by a humanistic orientation. The social reconstructionist orientation was least prevalent though still positively and moderately correlated with the three other curriculum orientations.

In a parallel study of teacher's conceptions of science curriculum, Cheung and Ng (2000) added a cognitive processes or skill orientation to the four previously identified major orientations. They developed a self-report instrument with a Likert-type response scale. Their results found that science teachers' orientations were predominantly cognitive processes oriented, though the other four orientations were not substantially weaker. This situation of many strongly held orientations is described by Cheung and Ng (2000, p. 367) as "complementary pluralism".

These five curriculum conceptions appear to have some element of the surface to deep continuum identified in the discussion of conceptions of learning. The order of conceptions from surface to deep seems to be technological, cognitive processes or skills, academic, humanistic, and social reconstructionistic. Thus, it is of interest to this study to evaluate whether teachers' conceptions of assessment are related in a meaningful way to their conceptions of curriculum.

Conceptions of Teaching

A number of independently developed models of teachers' conceptions of teaching (e.g., Gow & Kember, 1993; Pratt, 1992a; Samuelowicz & Bain, 1992; Trigwell & Prosser, 1997) have been compared (Kember, 1997) and show that three major approaches to teaching were found. The first is teacher-centred transmission of content (i.e., knowledge or information), while the second is a student-centred conceptual learning process. The complexity of teachers' mental realities, however, means that many teachers' conceptions of teaching lay between, as much as at either end of, the more surface-like first approach and the deeper second approach. The third approach is a bridging one that involves student and teacher interaction or apprenticeship. Kember (1997, p. 263) has argued that these conceptions are not hierarchical but rather "an ordered set of qualitatively differing conceptions" ranging from along the axis of teacher to student centred.

Thus, teachers have differing conceptions of teaching and this may make a difference to how they conceive of assessment. Gow and Kember (1993) argued that conceptions of teaching affect teaching methods used by teachers, the methods students use to learn, and the learning outcomes students achieve. In other words, teachers who conceive of teaching as being teacher-centred use a transmission of knowledge method (e.g., lecture) and their students acquire a surface reproduction of knowledge. Thus, it is argued that "the methods of teaching adopted, the learning tasks set, the assessment demands made and the workload specified are strongly influenced by the orientation to teaching" (Kember, 1997, p. 270). Ho, Watkins, and Kelly (2001) showed in a study of planned change of teacher conceptions of teaching that teaching practice improved promptly and student learning eventually improved when teachers adopted a more advanced conception of teaching. Jensen, Kauchak, and Rowley (2001) showed in a

study of four teacher trainees that the candidate with the most constructivist, deep

learning conception of teaching actually learned much more about teaching than the

candidate with the most behaviourist, transmission-oriented, surface learning

conception of teaching. Samuelowicz (1994) showed that two teachers with differing

conceptions of teaching had differing conceptions and practices of assessment. The

first teacher, who had a deep, student-oriented conception of teaching, emphasised

assessment as a means of improving teaching, providing feedback to students to

improve their learning, and as a means of making students accountable. This teacher

emphasised higher-order, problem solving, and decision-making processes in

assessment tasks. On the other hand, the teacher who was more transmission, teacher-

oriented conceived of assessment only as a means of forcing students to be accountable

for their learning and emphasised recall of knowledge in assessment tasks.

Conceptions of teaching are assumed to be dependent on learning contexts

(Marton, 1981) such that different conceptions may be prevalent depending on

"differences in stage of schooling, major subject area, curriculum, evaluation system,

social and cultural background, etc" (Gao & Watkins, 2002, p. 62). Prosser and

Trigwell (1999) developed a set of contingent principles for good classroom learning

and teaching. They suggested that teachers must be continuously aware of the students'

learning situations, the contextually dependent nature of teaching, the perceptions

students have of teaching technologies, the diversity of students in a class, and the need

to constantly evaluate and improve teaching.

Within the context of the three major conceptions of teaching (i.e., teacher-

centred, bridging, student-centred), there are models that identify more finely first-order

factors of teachers' conceptions of teaching. For example, Gao and Watkins (2002), in

a study of Chinese physics teachers, identified five conceptions designed along the

teacher-student centred continuum described by Kember (1997). These conceptions were identified as knowledge delivery, exam preparation, ability development, attitude promotion, and conduct guidance and were grouped into two approaches. The moulding approach involved the first two conceptions and the cultivating approach involved the last three conceptions. Gao and Watkins developed a 37-item questionnaire based on this model with over 700 secondary school physics teachers in People's Republic of China.

Another instrument, based on a similar model of teachers' conceptions of teaching has been developed in a series of multi-cultural studies (Pratt, 1992a; Pratt, 1992b; Hian, 1994; Pratt, 1997; Pratt & Associates, 1998; Pratt & Collins, 1998; Pratt & Collins, 2001). The *Teaching Perspectives Inventory* (TPI) attempts to measure beliefs about the nature of teaching by enquiring into teaching intentions, actions, and beliefs. There are five perspectives that parallel the range of teacher-centred, surface views to student-centred, deep views of the teaching-learning process. The most teacher-oriented conception, transmission, describes teachers who effectively communicate a well-defined and stable body of knowledge and skills to learners who must master that content. The other four conceptions are more student-oriented views of teaching. Apprenticeship assumes that the best learning happens when students work on authentic tasks in real settings of practice with learners gradually doing more of the work. The developmental perspective begins with the learners' prior knowledge and works towards restructuring how students think about that content through effective questioning and 'bridging' knowledge. The nurturing perspective respects students' self-concepts and self-efficacy in an effort to support student achievement by caring for the whole person not just the intellect. The social reform perspective views social and structural change as more important than individual learning and so teachers advocate

change in society as the purpose of teaching. A unique characteristic of the TPI among the various conceptions of teaching instruments is the place given to a social reform conception of education. Nevertheless, not only does this instrument have strong comparability to the student-teacher oriented axis, but it also contains parallels to the social reconstruction conception of curriculum developed by Cheung (2000) and thus was judged appropriate for use in this research.

Conceptions of Efficacy

Teacher efficacy refers to teachers' conviction or belief in their own ability to influence how well students learn or perform. Research into teacher efficacy has been shaped by two major traditions; Rotter's (1982) internal versus external locus of control and Bandura's (1989) self-efficacy (Tschannen-Moran, Woolfolk Hoy, & Hoy, 1998). Locus of control identifies whether control over outcomes resides within a person (internal) or in activities or circumstances outside the control of the individual (external). Self-efficacy, from Bandura's social cognitive theory, is belief or confidence in one's own ability to organise and take action in order to reach a goal. It is a conviction that one can successfully do what is necessary to achieve or produce a desired set of outcomes. High levels of self-efficacy impact positively on cognitive, motivational, selection, and affective processes individuals need to reach goals. It is also considered that self-efficacy is specific to a domain of goals rather than generalised to all possible situations or goals. The consequence of positive self-perception (i.e., self-efficacy) is effort to achieve goals, persistence when confronting obstacles, and resilience in face of adverse situations (Pajares, 1996).

Teachers' confidence in their own ability creates initiation of and persistence in courses of action that are capable of creating learning in students (Gibson & Dembo,

1984). Teachers' sense of their own efficacy as teachers has been related not only to positive teaching behaviours (e.g., lower stress levels, willingness to remain in teaching, and willingness to implement innovations), but also to increased student achievement, student self-efficacy, and motivation (Henson, Kogan, & Vacha-Haase, 2001; Tschannen-Moran et al., 1998).

Research has identified two major dimensions of teacher efficacy, though there is an on-going debate over the meaning of the dimensions (Tschannen-Moran et al., 1998). Gibson and Dembo (1984) identified, in their *Teacher Efficacy Scale*, the two dimensions as personal teaching or internal efficacy and general teaching or external efficacy. The former referred to a belief in one's own personal skills and abilities to bring about learning in students. Internal self-efficacy focused on statements related to personal responsibility in shaping student learning and behaviour. External efficacy referred to the belief that a teacher's ability to bring about learning is "limited by factors external to the teacher, such as the home environment, family background, and parental influences" (Gibson & Dembo, p. 574). This was associated with Bandura's outcome efficacy concept in which locus of control for student learning is attributed to either personal agency or external environmental factors. In other words, external efficacy refers to teachers' beliefs about whether students' learning outcomes could be controlled by the teacher efficacy or determined by external causal factors, such as home and family influences.

Guskey and Passaro (1994) questioned the meaning of these two dimensions in their revision of the *Teacher Efficacy Scale*. They rewrote items to reflect a personal internal agency ("I can") or a general occupational external agency ("teachers can"). Their research likewise identified two relatively unrelated ($r = -.24$) factors that fell into the internal and external pattern as proposed. However, because of the low inter-factor

correlation, they argued that these factors represent two separate beliefs rather than two ends of the one 'teacher efficacy' belief. In other words, teachers could have high personal internal agency beliefs ("I am an efficacious teacher), but simultaneously have low external environmental agency beliefs ("Teachers are not efficacious compared to student home and family factors").

Tschannen-Moran, et al. (1998, pp. 231-232) argued that the GTE or external belief factor is "a measure of optimism about the abilities of teachers in general to cope with adverse circumstances such as an unsupportive home environment or unmotivated students" and that it "taps teachers' tendencies to blame the home and the students for student failure". Other external factors, such as quality of curriculum resources, school leadership, school culture, and so on, may also affect external factor judgements but they are not captured in the present instruments. For example, Delandshere and Jones (1991) argued that their three mathematics teachers took the view that students' socio-economic conditions and students' fixed level of ability in the subject absolved the teachers from responsibility for student failure to achieve expected outcomes. Tschannen-Moran et al. (1998) further argued that internal factor statements about self-perception of teaching competence are a poor measure of teacher efficacy because the items mix present and future or hypothetical conditions, violating the assumption that self-efficacy is context specific. Thus, the present set of instruments available to measure teacher efficacy are limited and further empirical and theoretical work is needed to improve instrumentation of this construct. Nevertheless, use of the Guskey and Passaro (1994) instrument was justified for the time being.

Conceptions of Assessment

Researchers have suggested that there are at least three major purposes for assessment; improvement of teaching and learning, certification of students' learning, and accountability of schools and teachers (Heaton, 1975; Torrance & Pryor, 1998; Warren & Nisbet, 1999; Webb, 1992). These purposes can lead to different practices, and often there can be tensions between the purposes. For example, the tension between externally imposed accountability requirements and the improvement conception has created difficulties for New Zealand teachers (Dixon, 1999; M. Hill, 2000b). It is argued in this thesis that these purposes do exist in the minds of teachers. However, a fourth conception can be inferred from the literature on assessment purposes. Specifically, this is a rejection or disregard of assessment; in other words the treatment of assessment as irrelevant to the life and work of teachers and students. Thus, this thesis asserts that there may be four major conceptions of assessment held by teachers; (a) assessment is useful in improving teacher instruction and student learning by providing quality information for decision-making, (b) assessment is about accountability of students through certification processes, (c) teachers or schools are made accountable through internal or external evaluations, and (d) assessment is irrelevant or pernicious to the work of teachers and the life of students. It should be noted that the various conceptions might interact with each other. For example, it is possible that those who see assessment as irrelevant could also believe that improvement is the legitimate goal of teacher judgement but simply reject assessment as a legitimate means of reaching that goal. On the other hand, concern for improvement may associate strongly with school self-managed accountability but less strongly with the student certification view. Awareness of measurement error in assessment may lead

to an irrelevance view of assessment. The detailed characteristics of these conceptions are described below.

Assumptions About Assessment

Assessment is understood as any act of interpreting information about student performance, collected through any of a multitude of means or practices. Thus, assessment, according to the Department of Education in England (as cited in Gipps, Brown, McCallum & McAlister, 1995, p. 10-11) is a "general term enhancing all methods customarily used to appraise performance of an individual or a group. It may refer to a broad appraisal including many sources of evidence and many aspects of a pupil's knowledge, understanding, skills and attitudes; or to a particular occasion or instrument. An assessment instrument may be any method or procedure, formal or informal, for producing information about pupils: e.g., [sic] a written test paper, an interview schedule, a measurement task using equipment, a class quiz".

Further, it is assumed that assessment informs and influences decisions about "people, individually or in groups … to be informed and appropriate" (Thorndike, 1997, p. 6-7). Assessment provides information that allows valid and appropriate educational decisions to be made about such things as (a) what to teach students, (b) what order to teach important content, (c) which students to choose for certain teaching, (d) which programmes to place students into, (e) how to classify students, (f) identify if students are making progress, (g) determine if they are learning less than they should, (h) ascertain if they have learned enough, (i) ask how much learning is present, (j) ask how good learning is, (k) identify student readiness to learn, (l) what personal characteristics students have, (m) detect which students require specialist help, (n) determine if students have realistic self-awareness concerning their abilities or (o) how good the

curriculum is (Airasian, 1997; Cronbach, 1970; Gronlund & Linn, 1990; Mehrens & Lehmann, 1984; Thorndike, 1997).

It is also assumed that the quality of information obtained through assessment can effect the quality of educational decisions (Cronbach, 1970) and that the quality of data obtained can be evaluated through the collection of validation and reliability evidence. The quality of assessment information includes awareness of (a) any limitations of assessment information, including degree of inaccuracy in any measure, (b) dangers in over-reliance on any single measure, and (c) any unfair consequences for students (Cronbach, 1970; Hall, 2000; Linn, 2000; Popham, 2000; Thorndike, 1997). Unfortunately, many teacher-made or classroom assessments and intuitive judgements lack such quality indicators (McMillan, 2001b).

Conception 1: Improvement of Teaching and Learning

The major premise of this conception is that assessment informs the improvement of students' own learning and improves the quality of teaching. This improvement is associated with two important caveats; (a) assessment describes or diagnoses the nature of student achievement or performance and (b) the information provided by assessment is of sufficient quality to be considered valid, reliable, and accurate description of student performance. In this view, assessment is a range of techniques, including informal teacher-based intuitive judgement of capability as well as formal assessment tools, designed to identify the architecture of student learning, including impediments to learning and unexpected strengths.

Popham (2000, p. 1) eloquently and forcefully expressed the improvement conception; "if educational measurement doesn't lead to better education for students, then we shouldn't be doing it … you'll learn that the only reason educators ought to

assess students is in order to make more defensible educational decisions regarding those students. That's really why educators should be messing around with measurement-*to improve student learning*" [italics in original].

Scriven (1991, p. 20) defined improvement assessment or evaluation as "evaluation designed, done, and intended to support the process of improvement, and normally commissioned or done by, and delivered to, someone who can make improvements." This is most often an internal process carried out "by the staff of the originating institution" (Scriven, 1991, p. 22). Improvement assessment, therefore, is seen as positive, constructive, and acceptable since it seeks to analytically "'unpack' an overall grade in an illuminating way … by awarding separate ratings to several aspects of the work (originality, organization, mechanics, etc.)" (Scriven, 1991 p. 30).

It is noted that the improvement conception is associated with the term 'formative' in contrast to an accountability conception, which is associated with the term 'summative' (e.g., Dixon, 1999; M. Hill, 2000b). However Scriven (1991, p. 28) has argued that both formative and summative assessment require similar levels of rigour and technical quality otherwise "the accuracy of the mid-course corrections" is undermined. "Formative evaluation should *at least* provide a preview of a summative evaluation, since one of its most useful functions is to be an 'early warning system'" (Scriven, 1991, p. 28). Furthermore, within the improvement conception it is possible for both end-of-unit (summative) and mid-course (formative) assessment to be used to improve teaching and learning or to evaluate the quality of student learning or teacher instruction. In other words, assessment is assessment and only timing differences within the cycle of teaching, learning, assessing, planning, and reporting separate these supposedly different types of assessment. Thus, this conception is predicated on the

assumption that the duty of teachers is to improve the learning of their students and that assessment is a process for obtaining information to assist in this process.

In England the Offices of Standards in Education (as cited in Gipps, et al., 1995, p. 7) urged that, in order to improve student learning, assessment "should ensure that individual learning is more clearly targeted and that shortcomings are quickly identified and remedied". Furthermore, in order to achieve this broad and simultaneously rich understanding of student capabilities, assessment has to involve many parts, strategies, and techniques (Gronlund & Linn, 1990).

Thus, assessment is seen as a mechanism by which teachers can improve student learning and improve their own teaching. That assessment may contribute to the improvement of teaching has some grounding in the use of assessment to evaluate the effectiveness of teaching, check teacher planning, and judge whether students have learned what was taught (Gronlund & Linn, 1990). As Cronbach (1970, p. 24) put it "tests are equally important as an aid in evaluating treatments and maintaining their effectiveness. When the teacher gives an arithmetic test, he [sic] is testing his instruction as much as he is testing the students' effort and ability. If the results are poor, he should probably alter his method." Assessment experts agree that teachers can use assessment to evaluate and thus seek ways to improve their own work. However, there is little evidence that teachers actually consider their teaching as a possible object of change when students do poorly on assessments (Robinson, Phillips, & Timperley, in press).

Consequently, the improvement view tends to reject formal testing if it simply means more multiple-choice tests of lower order cognitive skills, such as recall or knowledge of discrete facts. This is so, notwithstanding, any legitimate place such assessment has in providing information about student's surface level understandings or

29

abilities. The challenge in the improvement conception is to identify the full range of student performance, including higher order thinking. In order to fully diagnose what students have learned, it is important that "test items provide definitions and criteria of successful learning and performing" (Firestone, 1998 p. 97). Thus, recent trends in educational assessment to implement authentic or performance-based assessment are predicated on the assumption that constructing a response to a realistic problem, such as writing an essay, showing how to solve a mathematical problem, or participating in a group multi-disciplinary social science simulation problem, requires students to show higher order cognitive skills such as application, analysis, synthesis, or evaluation (Firestone, 1998). Borko, Flory, and Cumbo (1993) reported that the adoption of performance assessment practices gave teachers more insight into student learning requirements.

If assessment tools are constructed to reflect classroom and curricular learning, then it is believed they "provide models of useful instructional strategies" (Firestone, 1998, p.97). As Hall (2000, p. 187) noted, "assessment must be coherent with course design, teaching and learning" for it to improve education. Integrating assessment with quality "curriculum and teaching practice" (Gipps, et al., 1995, p. 12) ensures teacher acceptance of improvement teacher assessment. Thus, "teachers are able to use the on-going assessment process formatively to encourage learning" (Butterfield, 1999, p.226) and assessment is legitimated if it results in action that successfully changes what students know or can do (Wiliam & Black, 1996). It has also been argued that improvement assessment is linked to a constructivist view of teaching; "a constructivist pedagogy is inherently concerned with the teacher's modelling of how individual pupils are thinking and understanding so that the next challenge, prompt, question or

information can lead the learner forward. The teacher's awareness of the understandings of pupils is predicated upon assessment" (Butterfield, 1999, p. 228).

The improvement view of assessment requires teachers to be actively involved in diagnosing and ascertaining what students have learned utilising a wide variety of evaluative techniques. To do this high quality information including, accurate and efficient description of a student's performance is required. New Zealand teachers reported using the standardised *Progressive Achievement Tests* most often for descriptive and diagnostic purposes of identifying students for further appraisal, comparing students to other students, grouping students for instruction, and planning instructional activities (Croft & Reid, 1991). This "depends upon a high level of understanding and involvement by teachers in assessment practices (Gipps, 1994, 158-166)." As a result, professional development of teachers in both pupil assessment and teaching is required so that they can implement improvement assessment. Butterfield advised, "steps must also be taken to increase their capacity to teach in new ways … teachers will need to understand short-term issues such as what it takes to score well on those tests. They may also need the deeper pedagogical content knowledge to help students learn the basic subjects at a more profound level" (1998, p. 98). It is argued, by the English Department for Education (as cited in Butterfield, 1999, p. 227), that responsibility for such professional development lies with schools that "will have a leading responsibility for training students to teach their specialist subjects, to assess pupils and to manage classes".

Having asserted that assessment ought to involve students in demonstrating higher order thinking, it is just as important to the improvement objective that the information obtained about those higher order skills be valid. Validity of assessment is obtained when the assessment method is consistent with the material or curriculum

being taught and if the results of the assessment are accurate. Accurate information is

ensured when there is good consistency between teacher judgements or between

assessment tools, avoidance of apparent subjectivity in scoring or grading, and in the

estimates of performance being relatively error-free (Gronlund & Linn, 1990;

Thorndike, 1997). Thus, a valid basis for describing student performance is established,

which in turn is the basis for helping students improve their own learning or guiding

teachers in changing their practice.

Indeed, the improvement process is enhanced when students, either through self-

assessment or peer assessment, are involved in the process of determining criteria for

evaluation and in conducting such assessment (Black & Wiliam, 1998; Crooks, 1988).

Students need to know how to independently evaluate their own work and become

aware of criteria for identifying possible enhancements in their own performance (Asch,

1976). Such an internalised locus of evaluation allows students to apply criteria in

independent situations where they can make knowledgeable judgments and select

directions for their own future endeavours (Asch, 1976). Sadler put the student

participation dimension of improvement well:

> For students to be able to improve, they must develop the capacity to
> monitor the quality of their own work during actual production. This
> in turn requires that students possess an appreciation of what high
> quality work is, that they have the evaluative skill necessary for them
> to compare with some objectivity the quality of what they are
> producing in relation to the higher standard, and that they develop a
> store of tactics or moves which can be drawn upon to modify their
> own work. (Sadler, 1989, p. 119).

In the improvement conception the purpose for assessing students' knowledge,

skill, performance, or understanding is to generate accurate information that leads to

valid changes in teaching practice or student learning such that improvement in student

achievement can be facilitated. This conception of assessment presumes that unless

evaluation leads to improvement, teachers are justified in conceiving assessment as

irrelevant or as something superficial related to checking up on teachers, schools, or students.

Conception 2: Accountability of Teachers and Schools

A second conception of assessment is that it can be used to account for a teacher's, a school's, or a system's use of society's resources. "Test scores give evidence about how well or badly … a school, or even a country is doing" (Firestone, 1998, p. 97). Rationales for the school accountability conception are two-fold; one rationale emphasises demonstrating publicly that schools and teachers deliver quality instruction, and the second emphasises improving the quality of instruction.

The first viewpoint insists that schools and teachers have to be able to demonstrate that they have are delivering the quality product that society is entitled to by virtue of funding the educational process (Crooks, 1990; Mehrens & Lehmann, 1984). The right of the state to ensure that quality is delivered for its funding has been extended in some jurisdictions so that "such accountability expresses itself as an increasing governmental prescription of curriculum and assessment, as a means of controlling and measuring the work of teachers" (Butterfield, 1999, p.225). As Smith, Heinecke, and Noble (1999, p. 183) reported, in the context of state-wide assessment in Arizona, "the 'problem', as defined by some … , was that schools were not accountable, efficient, or effective." Similarly, one reason for imposing the new accountability system in England, was the perception that "the lack of an examination or national system of testing at the end of primary school left this sector of schooling wide open to criticism of performance standards across the primary age range (5-11)" (Gipps, et al., 1995, p. 6). Indeed, for some, accountability "has come to mean the responsibility of a school (district, teacher, or student) to parents, taxpayers, or government (federal, state,

city, or district) to produce high achievement test scores (Smith & Fey, 2000). The

dilemma for many looking at the education industry is simple; "employees get paid to

come to work, but no relationship exists between what their labours produce and their

level of compensation. This disconnect between inputs and outcomes does not exist in

other professions and represents a fatal flaw in the current system" (Hershberg, 2002, p.

330).

The second viewpoint for making schools and teachers accountable for their

work is based in the idea that assessment for accountability focuses on improving the

quality of teaching and learning. One of the largest obstacles to teachers improving the

achievement or learning of their students is their own conception of what learning and

teaching are. The prolific use of multiple-choice tests is said to be indicative of a

lower-order skill and drill instruction based on a behaviourist psychology and pedagogy

(Noble & Smith, 1994). Educational reformers have argued (Resnick & Resnick,

1989), based on a cognitive-constructivist view of measurement driven reform (Noble

& Smith, 1994), that assessment could raise the quality of teaching and educational

achievement or standards. For such reformers it is important to change teaching

practice to sound cognitive and constructivist psychological and pedagogical principles

(i.e., learning is a process of construction, learning depends on knowledge, and learning

is situated in socio-cultural contexts) in order to improve student learning (Resnick &

Resnick, 1989). Since few teachers work on such a basis, it was believed that unless the

stakes were high enough most teachers would not change their practice. Thus, if

assessments exemplified sound learning and teaching principles (i.e., were holistic,

integrated, project-oriented, long-term, discovery-based, and social), then their

mandated use would force teachers to improve their teaching for the good of students.

Furthermore, if assessments were samples of truly valuable learning objectives or

standards, rather than just the easiest material to test with multiple-choice formats, then

not only would assessments improve teacher pedagogy and psychology but they would

also concentrate teaching content on important knowledge and learning. As Linn (2000,

p. 7) put it "the focusing of instruction on the general concepts and skills included in the

test may be in keeping with the belief that the test corresponds to instructionally

important objectives and may be considered acceptable, even desirable, practice".

"The purpose of implementing an assessment and accountability program in an

urban school district is to improve student learning of worthwhile content" (Porter &

Chester, 2002). The official justification in England (as cited in Gipps, et al., 1995, p.

5) for accountability assessment to improve education or raise standards is expressed as:

> A national curriculum backed by clear assessment arrangements will help to
> raise standards of attainment by (i) ensuring that all pupils study a broad and
> balanced range of subjects … (ii) setting clear objectives for what children
> over the full range of ability should be able to achieve … (iv) checking on
> progress towards those objectives and performance achieved at various
> stages, so that pupils can be stretched further when they are doing well and
> given more help when they are not.

The strategy in accountability assessment to improve schools and teachers is

multi-faceted. First, according to Linn (2000, p. 8) develop and use "ambitious content

standards as the basis of assessment and accountability" and second, set "demanding

performance standards" for all students, and third attach "high-stakes accountability

mechanisms for schools, teachers, and, sometime, students". Thus, those committed to

accountability believe assessment "will ipso facto raise standards" (Gipps, et al., 1995,

p. 8). It was this rationale that led to the development in England of a new national

curriculum and assessment regime in primary schooling (Gipps, et al., 1995).

The essence of any accountability system, whether intended to be punitive

towards teachers or intended to motivate improvement in instruction, is the use of

consequences or stakes for teachers, schools, or systems based on the results of student

assessment (Guthrie, 2002). "Pressure can normally come from stakes or sanctions, the administration of which depends on test scores. Stakes can be targeted at either students or educators and can take a variety of forms. Passing a test can be a requirement for graduation from school. The proportion of students achieving at a certain level can trigger consequences for educators ranging form merit pay to state takeover" (Firestone, 1998, p. 97). Accountability consequences can be either positive or negative, and tend to be high rather than low stakes. For example, consequences may include a student being held back (negative, high-stakes), a student being awarded scholarship for external examination results (positive, high-stakes), a teacher being given extra pay for increased student assessment results (positive, high-stakes), a school board or school being disestablished for continued poor student assessment scores (negative, high-stakes), or a school being publicly vilified in the media for having poor student assessment results compared to other schools (negative, high-stakes).

Accountability may have the goal of reducing teacher influence or control of education (i.e., perceived as teacher bashing); "for some, accountability … is a cudgel to beat up on the teachers unions and the educational establishment in general" (Hershberg, 2001, p. 329). The accountability view of assessment has significant impact on the work of teachers and school management or governance; "accountability programs took a variety of forms, but shared the common characteristic that they increased real or perceived stakes of results for teachers and educational administrators" (Linn, 2000, p. 7). This reality may lead teachers to see assessment as being largely antithetical to the welfare of teachers and thus provoke a strong adherence among practitioners to the irrelevance conception of assessment.

Whatever the reason for accountability evaluation of schools, teachers, and students, it is clear that student assessment is seen as the ideal means for delivering

accountability. Linn (2000) explained succinctly why assessment is given this large

responsibility.

> First, tests and assessments are relatively inexpensive. Compared to changes that involve increasing instructional time, reducing class size, attracting more able people to teaching, hiring teacher aides, or implementing programmatic changes that involve substantial professional development for teachers, assessment is cheap. Second, testing and assessment can be externally mandated. It is far easier to mandate testing and assessment requirements at the state or district level than it is to take actions that involve actual change in what happens inside the classroom. Third, testing and assessment changes can be rapidly implemented. Importantly, new test or assessment requirement can be implemented within the term of office of elected officials. Fourth, results are visible. Test results can be reported to the press. (Linn, 2000, p. 4)

A key approach to accountability assessment that raises standards is to involve

teachers as the assessors. Since teachers are well positioned to "monitor achievement,

to redirect student learning quickly and to implement a testing programme at the

appropriate points in the learning process performance" their involvement in

accountability will contribute to more valid conclusions (Hall, 2000, p. 189). "An

important aspect of the TGAT framework was that teachers' assessments would be

central to the system; teachers were to assess pupils' performance continuously using

their own informal methods and this assessment information would provide both

formative and diagnostic information to support teaching. … As Lawton (1992) pointed

out, a major achievement of the TGAT report was to make a significant change in

professional and public thinking in relation to ongoing assessment and the teacher's role

in this" (Gipps, et al., 1995, p. 12). It was hoped that involving teachers in assessment,

especially if the assessments corresponded "to instructionally important objectives"

(Linn, 2000, p. 7), "would transform large numbers of classrooms" (Firestone, 1998,

p.96) because "tests can serve useful persuasive and educational functions" (Firestone,

1998, p. 97). In Meyer's words (as cited in Linn, 2000, p. 12), "teaching to the test

could induce teachers and administrators to adopt new curriculums and teaching techniques much more rapidly than they otherwise would".

The positive consequences of accountability assessment have begun to surface. Cizek (2001) reported, among ten unintended consequences of high-stakes testing, improvements in the quality of teacher professional development, increased awareness and provision of accommodations for special needs students, increased assessment literacy among teachers, wider spread collection and use of achievement data, deepened understanding of tested curriculum content on the part of teachers, and increased student learning. High-quality assessments designed and aligned with curriculum reform have found, in Pittsburgh and Michigan, significant improvements in the achievement of low-SES districts and among non-white populations (Schoenfeld, 2002).

However, accountability practices have also been found to have negative impact on teachers and teaching (Cooper & Davies, 1993; Delandshere & Jones, 1999; Kohn, 1999; Noble & Smith, 1994; Smith & Rottenberg, 1991; Smith & Fey, 2000). The most obvious effect, from a teacher point of view, is teaching to the test (Firestone, 1998; Hall, 2000; Linn, 2000; Meyer, 1996). If there are high-stakes consequences attached to assessment results, then it is only rational for teachers and school administrators to do all within their power to ensure that the desired results are obtained. This has often been referred to as measurement driven reform (Noble & Smith, 1994). Meyer (1996, p. 140) has argued, "in a high-stakes accountability system, teachers and administrators are likely to exploit all avenues to improve measured performance. For example, teachers may "teach narrowly to the test." Such pressure has led "teachers to emphasize drill-based instruction, narrowing of content, and the regurgitation of facts even more

than they did normally … substantial time was lost in test preparation" (Firestone, 1998, p.98).

The public transparency element to the accountability conception requires that assessment results be public information. There is clear evidence that accountability assessment "can also have more direct consequences if they prompt public discontent or-as happens in England-if parents have the opportunity to choose their schools on the basis of past performance" (Firestone, 1998, p. 98). The 'league table' publication of assessment results creates winner and loser schools in the public arena. This can exacerbate rather than alleviate the problems of a low-scoring school in that parents and educators are driven away from such a school. In enrolment-funded provision arrangements this can have a compounding effect on a school's capability to address causes of low achievement (Fiske & Ladd, 2000) and can encourage schools and educators to resort to intensified teaching to the test.

The most serious problem with teaching to the test, beyond what it may do to a teacher's professional status or conception of instruction, is the uncertainty as to the meaning of any increase in test scores. "There is less consensus, however, as to … the extent to which imposed, legislated testing will genuinely raise education standards, or whether it will simply raise test scores at the expense of better education" (Gipps, et al., 1995, p. 9). In fact, research has shown that scores have increased in what is known as the Lake Wobegon effect, so named after Garrison Keiller's mythical town where all the students were above average. Cannell (1989) concluded that the misuse of standardized, nationally normed achievement tests, originally intended as instructional improvement tools, as accountability reporting tools gives students, parents, school systems, legislatures, and the press inflated and misleading reports on achievement levels. As Linn (2000, p. 7) summed it up "whatever the reason for the Lake Wobegon

effect, it is clear that the standardized test results that were widely reported as part of accountability systems in the 1980s were giving an inflated impression of student achievement". Koretz (2002, p. 323) identified "the single most threatening explanation, when you want to conclude that an accountability system is working, is the prospect of seriously inflated test scores".

Thus, it can be seen that accountability assessment for improving education is associated with several major claims; (a) assessment will help raise educational standards, (b) governments have to ensure that resources are being used well to reach those standards, (c) assessment will ensure that the mandated curriculum is taught, and (d) testing of students is an appropriate means to establish if teachers and schools have done all in their power to raise educational standards. However, both experts and practitioners contest the validity of these claims. Such contest by educators encourages acceptance and proliferation of an irrelevance view of assessment.

Conception 3: Accountability of Students

The premise of the third conception of assessment is that students are individually accountable for their learning through their performance on assessments. This is commonly seen in the various qualifications examinations that secondary age students participate in either for graduation or for entry selection to higher levels of educational opportunity. For example, students in New Zealand schools are made accountable in secondary school for their learning through their participation in high stakes qualification or certification assessment activities (whether those are conducted internally by teachers or by external agencies) in the final three years of schooling (e.g., School Certificate, Sixth Form Certificate, Unit Standards, National Certificate of Educational Achievement, or University Bursary examination). There are many and

usually significant consequences for individuals dependent on their performance on such assessments, including retention in a year or grade level, graduation, and tracking or streaming (Guthrie, 2002). Thus student accountability is largely about high stakes consequences such as graduation or selection or being publicly reported on as earning a certain grade, level, or score.

Another dimension to student accountability is students' participation in the setting of achievement goals that are marked by assessment events. While this may be more of primary school practice, it is certainly clear that motivating or requiring students to participate in self-managed learning that contributes toward being awarded a grade level or qualification level is a significant dimension of making students accountable for their learning. It is this type of continuous (Ojerinde, 2002), school-based (Keightly, 2002) or internal assessment that contributes towards certification that is objected to by many advocates of a more formative approach to assessment (Broadfoot, 2002; Gipps, 2002).

Students are participants in learning and teachers frequently obtain information about students through assessment processes, interpret that information in light of other students' achievement or in light of curriculum standards or individual's previous performance, and then report on students' cognitive and affective performances, achievement, attitude, and effort (McMillan, 2001a). Those reports, whether narrative or simple grades, make students accountable for their role in education and for their learning outcomes to themselves, their parents, and society. Reasons given by teachers in the United States for making students accountable through reports included ensuring that students' maintain motivation, effort, and attention at learning the material teachers considered important enough to present in class, rewarding cooperation, and penalising those who were inattentive, uncooperative, lazy, or unmotivated (Kahn, 2000). Note

41

that these are similar to the student control issues raised by Torrance and Pryor (1998). Another purpose in making students accountable for learning is to certify that they have attained the necessary standard for entry into higher levels of teaching, training, or education or employment opportunities (Guthrie, 2002).

Consequences attached to assessment of students are usually high-stakes when they are operated at a system level, for example passing a test in order to be promoted to the next grade level, failing to meet a cut-score on a diagnostic assessment that results in placement in a specialist remedial class or failing to meet the cut-score on a high-stakes test for entry into a prestigious enrichment class, or, at the secondary school level, passing examinations to receive nationally accredited qualifications that permit entry to tertiary institutions or the awarding of special scholarships or bursaries (Guthrie, 2002). As noted earlier, high-stakes assessment is a major component of the educational experience of adolescents in New Zealand schools with external qualifications assessments in the final three years of secondary schooling.

There are many opponents to over-reliance on high-stakes assessment for decisions that impact so significantly on the lives of individual students, grounded in both validity and reliability considerations (Kohn, 1999). High-stakes qualifications assessments may disadvantage students who were involved in a class that prioritised current or local events over material set down for the assessment (Crooks, 1990). It has been argued that high-stakes student accountability systems implemented through assessments (such as graduation criteria, placement in remedial programmes) are biased against certain population groups, specifically, low socio-economic and ethnic minority populations (English, 2002, Neill, 1997).

Teachers have concerns about the impact of assessment on students' affective dimensions; that is, "three of the mentors also stressed its affective role-making pupils

"feel wanted" or "feel good about themselves… and two of them stressed the importance of giving praise" (Butterfield, 1999, p. 233). Waltmann and Frisbie (1994) warned that such affective information must be reported separately from academic achievement and should not contaminate the assignment of achievement grades. Smith (1991) reported that teachers believed high-stakes testing to have an adverse emotional impact on young students, causing unwarranted worry and anxiety. Expert mathematics teachers have considered report card grading as an undesirable practice in contrast to assessment that informs teaching by checking on what students understand or have learnt (Philipp, Flores, Sowder, & Schappelle, 1994). Certainly, low achieving students "are likely to feel resentment, anxiety, lack of appropriate test-taking strategies, and decreasing motivation" when confronted with high-stakes assessments (Guthrie, 2002, p. 373).

However, because teachers and administrators are unable to accurately predict the effect an assessment evaluation will have on any individual student, it is difficult to assert that even gold star consequences are truly low-stakes. The individual student's experience of what are purportedly low-stakes consequences may in fact be traumatic and extremely high-stakes either through personality or family sanctioned consequences.

Conception 4: Irrelevance

The premise of the fourth conception of assessment is that assessment, usually understood as a formal, organised process of evaluating student performance, has no legitimate place within teaching and learning. Teachers' knowledge of students based on long relationship and their understanding of curriculum and pedagogy preclude the need to carry out any kind of assessment beyond the intuitive in-the-head process that

occurs automatically as teachers interact with students, what Airasian (1997) called 'sizing up' assessment. Assessment is rejected also because of its pernicious effects on teacher autonomy and professionalism and its distractive power from the real purpose of teaching, i.e., student learning. Teachers of English in England welcomed a new National Curriculum in the early 1990s but rejected the associated accountability assessments because the paper and pencil Key Stage assessments were considered inimical to the learning and teaching values espoused in the curriculum (Cooper & Davies, 1993).

Treating assessment as irrelevant is commonly connected to the claims (a) that assessment equates to testing, and, by corollary, that testing is bad for education, or (b) that assessment makes teachers, schools, and students accountable for their work. For example, Smith (1991) reported that some teachers whose schools received low scores in publicly released high-stakes testing programs questioned the validity of the tests. Negative attitudes to testing appear to increase as pressure to improve student scores increases and is especially noticeable among teachers who work with students from lower socio-economic homes (Herman & Golan, 1993). Research in England (Gipps et al., 1995) has identified and described teachers who resist 'assessment' by practising an intuitive, professional judgement of students' performances. "Their own assessments were intuitive and discursive, rather than against criteria, and often not written down; teachers found it hard to articulate their assessment practice" (Gipps, et al., 1995, p. 2). It is argued that this approach to assessment is widespread among teachers; "the general consensus [in Connor 1991] seemed to be that much of the assessment made by primary teachers was intuitive" (Gipps, et al., 1995, p. 11). These intuitive teachers were also detailed in Hill's (2000a) research among New Zealand primary teachers whom she described as head-note assessors. These teachers relied on their remembered and

interpreted observations of students' classroom participation in determining the status and requirements of a student. This is the type of assessment that Torrance and Pryor (1998) have advocated as 'true formative' assessment in contrast to the systematic collection of evidence for accountability to government requirements that they saw taking place in English classrooms.

The rationale for this conception of assessment lies partially in a student-centred focus on teaching. Since the whole student, including their self-concept, physical development, social relationships, etc., (not just their academic or intellectual growth), is seen as the core of student-centred teaching it is important for such teachers to take a much wider view of the student than is easily or more typically measured by an educational assessment. Indeed, Asch (1976) argued that teachers' student-centred beliefs, emphasising care for students' emotional and psychological well-being and morale, the valuing of individuality, prizing creative or divergent expression, and development of a positive, encouraging classroom atmosphere, seem to be factors in "a decline in the use of evaluative criteria" (p. 18). Torrance and Pryor (1998) described the student-centred culture as one where (a) students discover things for themselves, (b) students learn by concrete experience, (c) students' wider out of school experience is invoked in learning, (d) teachers elicit ideas from students rather than directly teach, (e) the teaching agenda is rarely made explicit to students, and (e) criteria for success are not made explicit to students. In such a worldview of education, it is easy to see how assessment may be conceived as evil or pernicious or at least as irrelevant.

It has been long established that New Zealand teachers are able to rank students in their own classes but are unable to reliably compare their own students with all other students in the nation (Crooks, 2002; Elley & Livingstone, 1972). Airasian (1997) in describing teachers' observation-based assessments, which he described as 'sizing up',

pointed out how ubiquitous and how untrustworthy they were for high stakes

consequences. Torrance and Pryor (1998) identified confounding variables in informal

teacher judgement assessment; (a) teachers seek to regulate pupil behaviour as much as

obtain information about student ability in their assessment interactions, (b) students

either resist or try to cooperate with teacher regulation conversations rather than provide

information about their real understanding, and (c) teachers often seek to boost the

confidence, motivation, or esteem of students rather than elicit information about their

ability. Though this is in effect a cautionary criticism of the irrelevance view of

assessment, it connects strongly to awareness of the inherent weaknesses of any one

approach to assessment and the need to use multiple methods of information. Further,

assessment expertise is required to be able to implement any conception of assessment

other than the irrelevance view.

In addition to any consequences that may be mandated because of student

assessment results, testing purportedly has the added capability of balancing the

subjectivity of teacher assessment (Hall, 2000). In the UK (as cited in Gipps, et al.,

1995, p. 15), the government intended to use external assessment to moderate the

distribution of teacher assessments, such that "the general aim would be to adjust the

overall teacher rating results to match the overall results of the national tests". This

checking of teacher work speaks of suspicion of teacher integrity and has contributed to

a teacher antipathy towards accountability assessment. However, formal assessment

tools (e.g., systematic testing) may be needed to complement teacher observation

because of the inadequacies of any one measure of student achievement (Gronlund &

Linn, 1990). Not only does assessment, whether formal or informal, contain error, but

assessment is "in reality, only representative samples of a target instructional domain of

skill, knowledge, or affect" (Popham, 2000, p. 85). Awareness of the error in assessment may contribute to the conception of assessment as irrelevant.

Dimensions such as effort, attitude, aptitude, behaviour, and previous progress all become important in evaluating students' learning. "Attitude is critical to effort and outcome and should be observed and comments recorded" (Limbrick, et al. 2000, p. 9). However, not everything that is important in education is necessarily appropriate for assessment in this conception. In discussion with teachers of Years 5 to 7, Limbrick, et al. (2000) found that "teachers for each year group indicated that it was not appropriate to assess attitudes in relation to the curriculum level" (Limbrick, et al. 2000, p. 9). McMillan (2001b) labelled facets such as effort, improvement, ability, participation and attention as academic enabling and reported that such components were widely used in secondary school teacher grading practices.

Another reason to avoid assessment is that it is perceived to be inherently subjective. Hall (2000 p. 178-179) argued that an impressionistic approach to assessment is appropriate because standards of achievement are held in the mind of the assessor, not in some external, objective curriculum. "Most educational standards … require subjective interpretation - the specified words are not enough… every marker or examiner carries his or her own interpretation of the standard into their assessment of student work… the marker will need to interpret their different performances to determine (judge) whether or not each has met the standard. The standard thus lies not only in the words that are specified, but in the mind of the marker, …" (Hall, 2000 p. 178-179).

This subjectivity, which Hall presented as a positive reason for rejecting assessment, also works in a negative sense to encourage students and teachers to reject assessment. Hawe (2000) demonstrated in a study of teacher-trainees that assessment

was conceived as being very much an artefact of individual teacher whim, personality, or caprice. Students reported that marking of assignments was inconsistent between instructors, standardised marking schedules were not used, and that grades were inconsistent with the effort they had made. As a result, assessment was frequently seen as a game that the student had to negotiate in order to get a fair result. The victims, that is, those receiving lower grades than expected, of such subjective assessment naturally were bitter, angry, and disillusioned. No doubt teachers fear inflicting such experiences on students and may use this perception as a basis for rejecting assessment as something pernicious in the relationship between teacher and student.

In cases where schools and teachers are required, perhaps through accountability provisions, to conduct formal assessment little use is made of the data (Robinson, Phillips, & Timperley, in press). When tests are required, teachers may prefer their own tests to nationally standardised or state-mandated tests for reasons of accessibility, proximity to and consonance with their own teaching (Hall, Carroll, & Comer, 1988). In other words, the data obtained through assessment is filed and forgotten. "Having carried out this level of testing, schools did little with the results beyond putting them into record books and using them to identify students for remedial help; little or no use was made of them for modifying teaching or curriculum – what we now call formative assessment" (Gipps, et al., 1995, p. 9). It is clear that for teachers a positive result of this irrelevance view is that "the burden of assessment is reduced, leaving them to focus more on teaching and learning" (Hall, 2000, p. 188).

Another dimension contributing to teacher reluctance to use or trust assessment is that fact that assessment is often used to control and regulate students. Butterfield (1999, p. 232) found "there was a significant focus across the majority of schools on class control and discipline, and 'assessment' was incorporated within this: questioning

would 'wake them up', a regular checking on homework would help to 'develop a work ethic'. In fact the marking of homework fulfilled "a policing function" (Butterfield, 1999, p. 232). Torrance and Pryor (1998) described the social regulation and control of student behaviour exercised by teachers through assessment interactions.

Another dimension contributing to teacher reluctance to use or trust assessment is the negative consequences assessments have on certain students, especially those disadvantaged by no fault of their own (e.g., minority students, new speakers of the language-medium of assessment, special needs students, those with reading difficulties, etc.). As Popham puts it:

> African Americans, Hispanic Americans, and most other minority groups
> have often suffered from educational testing practices that are
> unquestionably biased in favor of individuals from the majority culture.
> Educational tests have typically been written by white, middle-class
> Americans; tried out on white, middle-class students; and normed on
> white, middle-class students. Is it any wonder that youngsters from other
> ethnic groups or lower socio-economic strata would fare more poorly on
> such tests than children of the white, middle-class types who spawned
> those tests? (Popham, 2000, p. 44-45)

Since teachers believe that such assessment is anchored in a non-student centred view of the educational world then it follows that they "are likely to ignore assessments that model forms of teaching and conceptions of learning with which they disagree or that they do not understand" (Firestone, 1998, p. 98). Indeed, if assessment does not reflect what teachers believe they are employed to teach, then acceptance of assessment will be low; "generally, most teachers thought their district curricula were not well aligned with the state tests" (Firestone, 1998, p. 10).

Thus, there are a number of compelling reasons for teachers to conceive of assessment as something irrelevant to their everyday work or as actually pernicious towards their goals of enhancing individual student growth in the many social, affective, motivational, and psychological dimensions that surround educational achievement.

Assessment may unfairly impact on certain students, teachers may be forced to implement assessment but choose to ignore it, or assessment may be so inaccurate that it is unreliable.

Summary

The four, multi-faceted conceptions of assessment have support from a review of the literature about teachers' implicit theories. The improvement conception is anchored around four main dimensions; (a) assessment improves teachers' practice, (b) assessment improves students' learning, (c) assessment describes students' achievement, and (d) assessment provides valid information. The irrelevance conception is founded on three main dimensions; (a) assessment is bad for students, (b) assessment is used but ignored, and (c) assessment is inaccurate. The accountability conception is made up of two distinct but related consequence-based dimensions; (a) assessment makes schools, teachers, and systems accountable and (b) assessment makes students accountable for or certifies their learning. What is unknown at this stage is whether teachers hold these conceptions, how these various conceptions inter-relate, and how strongly teachers may hold each of these conceptions. In order to further examine teachers' conceptions of assessment, a model is required that can be tested. Together these four conceptions may constitute the basis for developing a model of teachers' conceptions of assessment.   Further, we need a model of teachers' conceptions that can show how those beliefs relate to other salient beliefs such as teaching, learning, or curriculum. It is the purpose of this thesis to develop such a model of teachers' conceptions of assessment.

Some models of teachers' conceptions of assessment have been developed that have been based on teachers' assessment practices or uses (e.g., Gipps, et al., 1995; M.

Hill, 2000b, Stamp, 1987). These models, based on types of assessment practices, relate to the model outlined in this chapter. Gipps, et al.'s (1995) model classified teachers by three major types of assessment (i.e., intuitives, evidence gatherers, and systematic planners). Intuitives emphasised professional, impressionistic, memory-reliant judgement processes of assessing students' performances intuitively without written records; a practice possibly related to the irrelevance conception. Evidence gatherers collected written evidence, usually at the end of units of work, to demonstrate students' progress relative to achievement objectives for the purpose of accountability; a practice possibly related to the school accountability conception. Systematic planners integrated systematic collection of multiple pieces of evidence of attainment of curriculum objectives with planned teaching for the purpose of shaping instruction; a practice possibly related to the improvement conception. Thus, the four conceptions of assessment developed in this review encapsulate Gipps et al.'s three types of assessment practice.

Hill's model, clearly related to the Gipps, et al. (1995) model, identified three reasonably distinct but related types of assessment use that could be used to classify teachers (i.e., unit assessment, head note assessment, and integrated systematic assessment). Unit assessors emphasised monitoring and recording, reasonably formally, students' progress on and achievement of curriculum learning objectives at the end of each unit of instruction; these teachers may conceive of assessment as primarily school or student accountability. Head note assessors, on the other hand, relied largely on their observation and memory of students' performance to occasionally record their impression of students' progression; these teachers may conceive of assessment as something irrelevant to their work of teaching. Integrated systematic assessors systematically planned and collected assessment data as part of focused teaching

activities for the joint purpose of recording progress and making improvements to teaching; these teachers seem to perceive assessment for improvement of teaching and student learning. Thus, the four conceptions of assessment described in this review encompass Hill's three types of assessment practice.

Stamp's (1987) model, developed with multivariate techniques, identified three major conceptions of assessment among pre-service teacher trainees in Australia (i.e., cater for need and progress of individual pupils, assessment blocks teachers' initiative, and a more traditional-academic summative examination). The first conception used assessment in a 'formative' way to identify individual student learning needs with the purpose of catering for those individual requirements; a set of beliefs quite similar to the improvement conception discussed above. The second conception reflected the view that teachers are required to conduct assessment but that assessment gets in the way of students' creativity and intuition, which are just as important as their academic development. This pupil-centred view is similar to the irrelevance conception. The third conception revolved around the use of tests and examinations to collect 'summative' information about students partly in order to motivate them to compete for more marks. This view is remarkably similar to the student accountability conception. Thus, the four conceptions of assessment reviewed in this chapter include Stamp's three types of assessment practice.

The three models of teachers' assessment practices outlined above suggest that teachers' practices fall mainly into one category. No evidence could be found for the relationship of practices to conceptions of assessment, for any interaction of the practices, or whether teachers mixed the conceptions in their practice. However, it is expected that teachers' conceptions of assessment interact with each other. For example, it is probable that teachers with an improvement conception of assessment

would treat as pernicious or simply ignore external accountability assessments for students, teachers, or schools. Furthermore, emphasis on assessment for improvement is expected to be inversely related to the irrelevance conception, in other words assessment for improvement is relevant. Garcia (1987) described a Spanish mathematics teacher who believed and practiced assessment for improvement, including seeking out information about the quality of his own teaching, and who at the same time begrudgingly implemented school-sanctioned student accountability assessment that he treated as irrelevant.

It is anticipated that the assessment for either school or student accountability conceptions would be positively related to the irrelevance conception; in other words these may be considered irrelevant. Nevertheless, the school accountability conception, in a self-managing professional environment, may be positively correlated with the improvement conception. It is likely that the two accountability conceptions will be correlated.

Additionally, though not surprising given that the nature of teachers' conceptions of assessment is unknown, it is unknown whether certain characteristics of teachers or schools influence teachers' conceptions of assessment. For example, the kinds of assessment methods teachers associate with the term 'assessment' may incline teachers to different assessment conceptions. The kinds of assessment methods teachers actually use and the length and type of assessment literacy training may also correlate with certain assessment conceptions. A teacher's role in a school, his or her length of experience, or gender may influence the conception of assessment held. The kind of school in which the teacher works (e.g., school size, or school socio-economic status) may also influence teachers' conceptions of assessment.

Conclusions

This review has highlighted the multifaceted nature of teachers' instructional conceptions. While the major concentration in the studies documented in this thesis has been on developing a model of assessment, the review has also identified meaningful dimensions of other instructional conceptions, such as learning, curriculum, teaching, and efficacy. This chapter has developed the basis for a model of teachers' conceptions of assessment by discussing four main conceptions of assessment. Three of those conceptions are based on significant purposes and the fourth is a rejection of some or all of those purposes. The literature on significant instructional beliefs (i.e., learning, curriculum, teaching, and teacher efficacy) has been reviewed in preparation for research into how teachers' conceptions of assessment relate to these pertinent beliefs.

Thus, it is possible to imagine through the following scenario how teachers' conceptions of assessment, learning, curriculum, teaching, and efficacy interrelate.

Five teachers (A-E) stand at the notice board in a school staff room and read a new notice that reads, "*Ministry of Education releases new assessment tool for literacy and numeracy. Contact the Regional Office for your copy.*"

A: See! All they're interested in is checking up on us. How can they keep using tests to decide if we're good teachers or not? What's the union doing to protect us?

B: Nah, tests are just a way to make the rich, white families feel good about paying heaps for the extra fees that their schools can charge. Our kids haven't got a hope of doing well. It's just part of a plot to keep minority students down.

C: That might be, but you know what to do, don't you? If they make you use it, just do it, write the scores down and forget about it and carry on doing what you always do. After all we're good teachers, we know what our kids are like and what they need. We don't need any tests to help us do a good job!

D: I'm not so sure about that. I've seen the trial stuff when our kids did it last year. The kids in my class really enjoyed them – it made them work a little harder and feel good about themselves. I think this kind of assessment might just motivate our kids.

E: Well, I've seen them too and I think the reports will help us do our jobs better. There are all kinds of descriptive information in them about what achievement objectives kids need to work on, what their strengths are, and what they've already mastered. It gives you all sorts of good ideas about where to start and who needs what.

A: But how can you trust a short test? All those fancy tests are full of error – you can never trust the results. Plus they're biased against our kids – the content's bound to be stuff that our kids won't know anything about. And the ESOL kids won't have a chance with the maths – there's just too much reading!

B: Well, even if they're really good informative tests, you know the boss will only want to use them to make the school look good, and find out who doesn't get the performance bonus at the end of the year. If your class grades go up, you get the money. It's so unfair because anyone knows that our kids start from so far behind they can never gain a lot.

C: Don't be silly! You teach them what's on the test, coach them well, and they can learn this stuff. They don't exactly ask questions that require rocket science ability on this kind of tests. Just start teaching them early on how to pass tests and they'll do fine and you'll get your money. Besides, the kids like you so they'll do their best for you and that's all that matters.

E: Gee, no wonder you get through your marking so quickly! All you do is write the numbers down, and ignore them? I spend a lot of time thinking about what the scores mean and how they relate to the stuff I've decided my kids need to learn this year. Then I use the information to change what I'm doing in my next bit of work; like maybe I find out that I need to go over something again for some of my kids.

D: I'm with you. We're all good teachers, we know we make a difference to our kids, and new assessments can show that we really are making a difference. But of course the tools have to be trustworthy and provide accurate information otherwise we may as well just carry on using our own quizzes instead of theirs.

B: Well, I hope you're right. I'm just worried that what they want to do with the test information is label the kids who struggle and put them into low-end remedial programmes while the so-called bright kids get special programmes. I think it's really unfair how tests are used to decide who gets promoted or certificated when we teachers really know what the kids can do. Even the kids know better, and all you have to do is ask them instead of using a test, for goodness sake.

In this hypothetical conversation, four major conceptions of or belief patterns about assessment are exhibited; (1) assessment is related to improvement of student learning and teachers' teaching (teachers D and E), (2) assessment certifies that students have learned or met standards (teacher B), (3) assessment evaluates the quality of schools and teachers (teacher A), and (4) assessment is irrelevant to the work of

55

teachers (teacher C). These conceptions are the filters through which teachers interpret and experience state sponsored or school-wide policies and practices related to assessment. Furthermore, the conversation is intended to point to an association among teachers' conceptions of assessment, teaching, curriculum, learning, and their effectiveness (e.g., Teacher E associates assessment with curriculum and instruction, Teacher C connects assessment to instruction and teacher efficacy, while Teacher D connects assessment to student learning).

Thus, it seems appropriate to investigate how teachers conceive of assessment and relate those conceptions to their conceptions of learning, teaching, curriculum, and teacher efficacy to get a fuller understanding of what assessment means to teachers. It is possible that the surface-deep continuum may be a useful organising schema for these various conceptions. If that continuum is of use then, surface views of learning should correlate with teacher-centred views of teaching, accountability and irrelevance conceptions of assessment, external views of teacher efficacy, and technological conceptions of curriculum. In contrast, deep views of learning should correlate with humanistic views of curriculum, student-centred perspectives of teaching, internal views of teacher efficacy, and improvement conceptions of assessment. On the other hand, there may be some other organising schema at work by which teachers' conceptions are related and given meaning.

In order to test these predictions and determine whether the demographic or individual characteristics effect teachers' conceptions of assessment a series of questions were posed. Further, in order to test predictions about how teachers' conceptions of assessment, teaching, learning, curriculum, and teacher efficacy are related further questions were posed.

1. Do New Zealand teachers exhibit views about learning and assessment that can be mapped to the conceptions identified in the literature review? (Studies 1 and 2)

2. Can a self-report attitude inventory about teachers' conceptions of assessment be developed? (Studies 3, 4, and 5)

3. Do teachers' conceptions of assessment fall into a four-facet model? (Studies 1, 2, 3, 4, and 5)

4. Are teachers' conceptions of assessment multi-faceted and hierarchical? (Study 5)

5. How do teachers' conceptions of assessment relate to each other? (Study 5)

6. What assessment methods do teachers associate with the term 'assessment' and what relationship is there between those association and conceptions of assessment? (Study 6)

7. What assessment practices do teachers use and what relationship is there between uses and conceptions of assessment? (Study 6)

8. What relationship is there between individual teacher characteristics (i.e., gender, role, experience, assessment training) and teachers' conceptions of assessment? (Study 6)

9. What relationship is there between school characteristics (i.e., size, SES) and teachers' conceptions of assessment? (Study 6)

10. How do teachers' conceptions of assessment relate to their conceptions of teaching, learning, curriculum, and teacher efficacy? (Study 6)

11. Is there a meaningful structure that relates teachers' conceptions of assessment, curriculum, teaching, learning, and efficacy?? (Study 6)

Thus this thesis, through a series of six studies proposes to answer eleven questions related to teachers' conceptions of assessment and how those conceptions relate to teachers' conceptions of the instructional constructs of teaching, learning, curriculum, and teacher efficacy. It is expected that large scale sampling of teachers' attitudes with a Likert-type response scale analysed with sophisticated correlational techniques will identify a meaningful structure in teachers' conceptions.

Study 1 will examine the instructional beliefs of secondary teachers from seven schools. Study 2 will investigate the conceptions of a large sample of primary school teachers. Studies 3 to 5 will follow the development of an instrument to measure teachers' conceptions of assessment towards a four-factor model. Study 6 brings together teachers' instructional conceptions into a second four-factor model.

CHAPTER II. EXPLORING TEACHERS' INSTRUCTIONAL CONCEPTIONS

This chapter reports two studies of teachers' conceptions of various instructional beliefs, including learning, curriculum, teaching, efficacy, and assessment. These studies were carried out with the goal of adding to what is known about how teachers think about these instructional constructs and how they might describe the relationship of those constructs to assessment. These studies attempt to provide evidence about whether the conceptions teachers have about assessment could be mapped to the conceptions identified in the literature review. Furthermore, these studies generated initial data on the conceptions teachers have about learning, teaching, curriculum, and teacher efficacy and how those conceptions related to each other.

Study 1 was an interview and survey exploration of secondary teachers' understanding of learning and assessment, while Study 2 was an analysis of free response opinions about assessment from a large survey of primary teachers. Study 2 uses largely qualitative approaches to data, while Study 1, though primarily dependent on interpretive methods, also made use of correlational analysis of teacher responses. The quantitative methods used in Study 1 permitted the testing of a six-point Likert-type agreement response scale as a means of eliciting responses from participants.

Study 1, specifically, examined how secondary school teachers' learning conceptions related to their conceptions of assessment, teaching, and teaching efficacy. It was expected that teachers' instructional conceptions and practices would be affected by the qualifications assessment system. Study 2, on the other hand, was designed to elicit more detailed understanding of primary school teachers' conceptions of assessment. It was expected that teachers would emphasise improvement conceptions,

and that they would also express significant concerns about assessment that would be linked to the irrelevance conception of assessment. This study offered the opportunity to examine primary teachers' beliefs to determine whether they were significantly different to those of secondary teachers.

Study 1: In-Service Secondary Teachers' Conceptions of Learning

In the New Zealand Curriculum Framework (NZCF) (Ministry of Education, 1993), explicit attention to learning is paid in the Essential Skills section under the heading Work and Study Skills. Students are required to (Ministry of Education, 1993, p. 20) (a) work effectively by themselves and in groups, (b) develop sound work habits, (c) take increasing levels of responsibility for their work, (d) choose realistic career paths, (e) take increasing responsibility for their own learning by building on their individual backgrounds, experiences, and styles, and (f) develop a desire and the skill for lifelong learning. These are further developed under the terms self-management and competitive Skills in the NZCF. "Students will: (a) set, evaluate, and achieve realistic personal goals; (b) manage time effectively; (c) show initiative, commitment, perseverance; (d) develop constructive approaches to success and failure; (e) develop the skills of self-appraisal; and (f) achieve self-discipline" (Ministry of Education, 1993, p. 19). These skills are typically aspects of self-regulated learning.

Thus, one of the major educational thrusts of the NZCF essential skills is that of individual, self-regulated, lifelong learning. Individual learning is a self-directed autonomous process wherein the student reflectively integrates the subject matter being studied (Chang & Simpson, 1997). Self-regulation of learning, explicit in all these statements about learning, concerns a "self-directive *process* through which learners

transform their mental abilities into task-related academic skills" [italics in original] (Zimmerman, 2001). The NZCF expects students to take responsibility for their learning, to be motivated, and to have self-awareness about what they are doing as well as have the skills needed to learn. In other words, the skills required of students have volitional, metacognitive, motivational, and self-regulatory characteristics (Masui & De Corte, 1999).

Study skills as understood by the NZCF are learning skills used in a myriad of events where an individual's mastery of learning must be demonstrated; e.g. the ubiquitous driving test, sports competition, workplace evaluation of information communication technology skills, and so on. Both students and teachers, although they are not intended just for academic assessments, commonly encounter study skills during examination preparation. The point of learning is that it must be demonstrated in some sort of 'assessment'; whether that be an essay examination, an in-class test, a practical performance such as completing a hill-start or parallel parking, or making economically effective consumer decisions. Learning is demonstrated publicly so that the degree and quality of that learning is known and recognised. This requirement that teachers are required to enable students to learn is partnered by the student and community expectation that teachers will ensure that students succeed in the assessments associated with the courses they teach.

In the context of Year 11 courses for high stakes qualifications, it was decided, as part of a global examination of how teachers conceive of assessment, to investigate what secondary teachers believed about learning and the effect, if any, that the assessment regime might have on learning beliefs. Specifically, the study addressed (a) what teachers believed learning to be, (b) what role teachers assigned to metacognition and motivation, (c) what teachers taught their students to do in order to learn for the

assessments of their subject, and (d) whether those conceptions and practices were the same for teachers of English, mathematics, and science at Year 11.

For the purposes of this study, study skills were understood to be a diverse, flexible, self-regulated set of strategies that learners deliberately use to learn some content or process such that the ideas, information, or material learned could be put to use or demonstrated under publicly observable conditions. The teacher's role in this process is understood to involve not only teaching students the ideas, information, or material but also to make students aware of a range of intentions and conceptions about learning and to train students in using a wide range of surface and deep learning strategies and tactics. There is little current evidence available on the state of New Zealand schoolteachers' awareness, practices, or conceptions about teaching study skills.

A sample of six secondary schools in the greater Wellington region, that represented a range of socio-economic statuses (SES), school roll sizes, and geographic locations, agreed to participate. About 80 teachers from the English, mathematics, and science departments of each school completed a questionnaire and 1 teacher per department per school was interviewed.

Both quantitative and qualitative methods were used. Bempechat and Drago-Severson (1999), in their review of cross-cultural motivation achievement research, explained why.

> "[Q]ualitative research allows us to illuminate the varied ways in which students, parents, and teachers ascribe meaning to academic achievement and educational experiences, quantitative analyses allow for generalizations about a particular group to a larger population of students." (Bempechat and Drago-Severson, 1999, p. 307)

Traditional methods of data collection, such as questionnaires and surveys, do not allow access to "free floating and far-reaching narratives" (Bempechat, et al., 1999,

p. 304). The qualitative methods appropriate for this study included focus-group interviews of students and semi-structured individual interviews with teachers. This multi-method approach permitted a deeper understanding of what students and teachers thought was required for successful studying and achievement. This approach had been used in a study of 88 Australian Year 11 students (equivalent to Year 12 in New Zealand) where a questionnaire was followed up by semi-structured interviews of a sample of students and with the teachers of the selected classes (Campbell, Brownlee, & Smith, 1996).

Instruments

Two recent reviews (Dubois & Kiewra, 1996; Hattie, Biggs, & Purdie, 1996) indicated that comprehensive and accurate models of studying still do not exist, notwithstanding Rohwer's (1984) much earlier call for the development of a psychology of studying. One of the most effective ways of testing a model is to test it through the development of a data collection instrument. New Zealand teachers have had the *Study Habits Evaluation and Instruction Kit* (SHEIK) (Jackson, Reid, & Croft, 1979) student self-report inventory available as a means of obtaining data about student study skills. The major criticisms of this instrument relate to the model on which SHEIK was developed. The model did not (a) apply study skills to actual course content, (b) discuss metacognitive approaches to studying, (c) link item construction to any theoretically or empirically based model of studying, (d) discuss the contingencies of study skill strategy selection and orchestration, (e) advance a wide range of study skills strategies, including spatial representation of learning and connection to prior knowledge, and (f) provide sufficient practice opportunities for effective learning

(Christenson, 1992; Kiewra. 1992). For these reasons, it was decided not to use or adapt for use the SHEIK with the teachers in this study.

Other overseas instruments were reviewed to establish whether they could be used or adapted for use with teachers. Entwistle's *Approaches and Study Skills Inventory for Students* (*Assist*) (Entwistle, n.d.) has three main approaches: deep, strategic, & surface-apathetic. Schmeck's cognitive psychology, information processing model of studying as expressed in the *Inventory of Learning Processes* (Schmeck, Ribich, & Ramanaiah, 1977; Schmeck, 2000) has four main factors: deep processing, methodical study, fact retention, and elaborative processing. Biggs' (1987b) widely used *Learning Process Questionnaire* has three approaches (each of which has strategy and motivational dimensions) to learning: achieving, surface, and deep. The *Motivated Strategies for Learning Questionnaire* (Pintrich, Smith, Garcia, & McKeachie, 1991), among its fifteen scales, measures strategy use, metacognition, and self-regulation. Weinstein and Palmer's (1990) *Learning and Study Strategies Inventory* for secondary students (LASSI-HS) assesses metacognition, motivational attributions, and study strategies. Vermunt (Vermetten, Vermunt, & Lodewijks, 1999) has developed, in Holland, an *Inventory of Learning Styles* that has four interrelated domains of learning; i.e. cognitive processing, metacognitive regulation, conceptions of learning, and learning motives.

After review of the various instruments, it was decided to use a short six statement section from Entwistle's *Assist* questionnaire because it was aimed directly at the surface-deep continuum of learning conceptions initially advanced by Marton and Saljo (1976). Furthermore, it was decided to allow teachers the opportunity to freely supply their own definitions of studying by asking them what they understood studying to be.

The issue of response format is not unproblematic (Gable & Wolf, 1993). The instruments from which the research instrument was derived use a variety of qualitative and quantitative response formats. There are clearly problems associated with how people respond to questions about both how often something happens and what their attitude is toward something (Sudman, Bradburn, & Schwarz, 1996; Schwarz, 1999; Schacter, 1999). Since the instrument was used early in the school year when teachers of Year 11 students would most likely have done very little teaching about studying, the use of quantitative frequency response formats was considered problematic. It was decided with the learning conceptions statements to use an agreement response scale because the statements were not conducive to frequency measurement and because they require a more holistic, opinion or attitude, response. A parallel study on student conceptions of studying found that the frequency response format produced less fitting responses on the whole than the agreement response format (Brown, in review).

A further complication with the response format is the nature and number of response anchors. An odd number of anchors, with an 'undecided' category, creates uncertainty as to whether an undecided response means not understanding, being neutral, being indifferent, or being acquiescent. Thus, an even number of response categories was used. Furthermore, since it was expected that participants would respond positively to these items, it was decided to use four positive and two negative terms to increase the variance of responses. In contrast to the common usage of negatively worded statements within questionnaires to reduce response sets (Gable & Wolf, 1993), Brown (in review) found, in a study of student conceptions of studying, that all the negatively worded statements had poor fit to an underlying single parameter item response theory model. Thus, the questionnaire items on learning conceptions were all worded positively.

The learning conceptions questions were designed to explore the deep, and surface approaches identified in the literature. Table 1 shows the number of items broken down by approach. Three statements were surface oriented because they emphasised using, getting, or remembering information or details. The other three items were deep oriented because of their emphasis on personal development, different perception, and understanding for itself.

Table 1. Learning Conceptions Items by Approach

Approaches	
Surface	Deep
Learning is being able to use the information I've got	Learning is understanding new material for myself
Learning is building up knowledge by getting facts and information	Learning is developing as a person
Learning is making sure I remember things well	Learning is seeing things in a different and more meaningful way

The interviews were audio-recorded and a research assistant assisted with written notes. The notes were used to identify trends and patterns of thinking, belief, or understanding about learning that were then classified into the surface, achieving, and deep categories. Teacher written definitions were analysed for conceptions about learning processes (i.e., strategies or methods) and conceptions about their understanding of learning intentions (i.e., motivations or goals). The teacher definitions were classified into surface or deep approaches. Given the correlational modelling techniques planned for the Likert-type statements, this level of qualitative data collection was deemed satisfactory for triangulation purposes.

Gow & Kember (1993) used a questionnaire and interview procedure in their study of teaching conceptions of polytechnic lecturers. In addition to the questionnaire and open-ended definition of studying, semi-structured interviews were used to elicit from teachers data about their teaching goals and approaches, their conceptions about

teaching and teachers, their understanding of assessment, and the attributes of a good

student (Table 2).

Table 2. Interview Schedule for Teachers' Conceptions about Curriculum, Teaching, Assessment, and Students

Topic Area and Questions
Conceptions about Curriculum
What subject and qualification level are you teaching at Year 11?
What are the goals of your subject at this level?
Why are the goals sometimes not achieved?
Conceptions about Teaching
What teaching approach do you take in this subject at this level?
Have you made any recent changes to your teaching approach in this subject?
What is your view of teaching? Would it be different if you were teaching students at a higher year level or a different level at Year 11?
What are the attributes of a good teacher in subject at this level?
Assessment Conceptions
What level of understanding versus recall of this subject do the assessments require?
Can the assessments be passed just by memory, revision, or practice?
Do the highest scoring students get their grades from deep understanding or through surface memory?
What kind of studying or learning do students need to engage in to ensure do well in assessments of this subject?
Students
What are the attributes of a good student in this subject?

The interview data was analysed to determine the common patterns and their

frequencies of expression. Since the number of participants was small, the

generalisability of this information is weak. Nevertheless, useful insights were gained

by over-viewing the main trends.

Analysis

The Likert-type response questionnaire items about learning conceptions were

scored by assigning numeric values to each response (i.e., 1=strongly disagree,

2=disagree, 3=slightly agree, 4=moderately agree, 5=mostly agree, 6=strongly agree),

on the assumption that each response represented a score on an underlying continuum

of agreement attitudes towards the concept. Thus, high scores represent strong agreement with a statement. In the questionnaire, participants who provided more than one response per item had the lower of the two responses taken if the two responses were side-by-side. If their multiple responses were further apart then the response was classified as missing data. It should be noted that the response format and scoring procedure used in this study with the Entwistle statements about conceptions of learning were used in Studies 3 through 6 in this thesis for all work on conceptions of assessment, curriculum, teaching, and teacher efficacy.

A number of sophisticated correlational analyses were conducted on the Likert-type data. Exploratory factor analysis (EFA) was used to determine whether responses to statements indicate a relationship of statements to a common factor while confirmatory factor analysis (CFA) was used to test whether the proposed set of relationships can be found in a set of responses (Gable and Wolf, 1993). Exploratory factor analysis was undertaken using maximum likelihood methods, and oblimin oblique rotation to determine the factor structure of teacher responses to the conceptions questionnaire. Where a theoretical structure has been proposed and when sample sizes are sufficient it is appropriate to use CFA techniques to test the fit of the proposed model to the data.

Participants

The number of teachers participating in the questionnaire by school size, subject, and school socio-economic status are identified in Table 3. Nearly three-quarters of teachers (70%) worked in large (roll >750) schools, while approximately equal numbers of teachers worked in each of the three departments of English, mathematics, and science. Almost a quarter (23%) of teachers worked in low SES schools, nearly a third

(30%) worked in middle SES schools, and the balance (47%) in high SES schools.

Nearly all of the teachers (73) taught School Certificate examination preparation

classes, while eight of the teachers taught alternative, non-School Certificate, courses

for students in Year 11 not intending to take the School Certificate examination.

Table 3. Teacher Questionnaire Participants by School Size, School SES, and Subject

	School Size					
	Large (roll >750)			Small (roll <751)		
	Subject			Subject		
School SES	English	Mathematics	Science	English	Mathematics	Science
Low	8	6	5	-	-	-
Middle	6	6	4	3	2	3
High	5	9	8	4*	6*	6*
Total	19	21	17	7	8	9

Note. *Two single sex schools contributed to this number.

Almost two-thirds of respondents had taught for ten or more years. Thus, the

sample is relatively more experienced than the full population of teachers of whom only

49% had taught ten or more years (Sturrock, 1999). Nearly 40% of respondents had

been in their schools up to only two years. In summary, the data obtained from the

questionnaire came from a group of Year 11 teachers who had (a) been teaching for less

than the national average, (b) been in their schools a relatively short time, (c) taught

(among other years) at least one Year 11 classes all their teaching careers, and (d)

subject related degrees. The majority of respondents had taught Year 11 classes the

whole of their teaching careers. Only 2 respondents had no degree, though it appears

that all the teachers of the three core subjects had degrees.

From the group of teachers who completed the questionnaire, 18 were selected

for interviews that were conducted in May 2000. Teachers were selected to ensure a

balance of males and females and so that at least one teacher of an alternative class for

each subject was included. Table 4 shows the gender, decile, and level breakdown of

interviewed teachers. Eight of the teachers were male and ten female, while three

taught alternative level courses (one per subject). Three were from a low decile school, six from two mid-decile schools and nine from three high decile schools. Six were from each of the three core subjects of English, mathematics, and science.

Table 4. Interviewed Teachers by Gender, Decile, and Teaching Level

	English				Mathematics				Science			
	Male		Female		Male		Female		Male		Female	
Decile	SC	Alt	SC	Alt	SC	Alt	SC	Alt	SC	Alt	SC	Alt
Low	1				1							1
Mid		1	1		1	1					2	
High	1		2		1		2		2		1	
Totals by Level	2	1	3		2	1	3		2		3	1
Totals by Gender	3		3		3		3		2		4	

As reported in Table 2, 81 teachers responded to the questionnaire. Clearly, some teachers teach more than one of the core subjects, since 42 indicated that they taught mathematics, 27 taught science, and 26 taught English. From this breakdown, it was further determined that 26 mathematics, 24 science, and 23 English teachers were teaching classes at the Year 11 level.

Results

Data from both the questionnaire and interview are reported together as they relate to the various conceptions of interest; specifically, conceptions about learning, teaching, assessment, and efficacy.

Conceptions About Learning

Only 14 teachers provided free-response comments about their understandings of learning in the questionnaire. Of those, 12 teachers gave definitions that focused on deep understanding intentions and methods, while only two were oriented to increasing student achievement. Thus, the responses teachers provided to the six learning conceptions statements gave the most insight into teachers' thinking and were used as the primary source of data.

Teachers tended to strongly agree with all the statements about the nature of learning except for the two surface questions, which emphasised using information and remembering things well respectively. An exploratory maximum likelihood factor analysis with oblimin rotation found two correlated ($r = .367$) factors (Table 5). The surface factor contained the 'remember information' and 'get information' statements, while the 'use of information' statement loaded equally on both surface and deep factors. The deep factor contained statements consistent with the deep categories identified by Marton and Saljo (1976). The poor fit of the 'use of information' statement with either the surface or deep factor was unexpected and suggests that use of information could be seen as either a deep or surface conception. Further research is needed to clarify this lack of fit.

Table 5. Teacher Conceptions of Learning Factor Structure

	Conception of Learning	
Learning Statements	Surface	Deep
Making sure I remember things well	**1.05**	-.18
Building up knowledge by getting facts and information	**.42**	.07
Being able to use the information I've got	.18	.16
Seeing things in a different and more meaningful way	-.02	**.88**
Understanding new material for myself	-.04	**.62**
Developing as a person	.03	**.55**

The mean score for each question and deep and surface scale scores, based on the factor pattern structure, was calculated (Table 6). Data from a parallel study on student conceptions about learning that involved approximately 700 students (details reported in Brown, 2002) is included in Table 6 to show important differences between student and teacher responses. The standardised effect size difference between teachers' and students' mean factor scores was calculated by subtracting the student mean from the teacher mean score and dividing by the pooled standard deviation weighted by sample size.

Table 6. Mean Learning Conceptions of Students and Teachers

Learning Belief	Students	Teachers	Effect-Size
Surface			
Building up knowledge by getting facts and information	5.07	4.95	
Making sure I remember things well	4.45	4.23	
Mean (SD)	*4.76 (1.016)*	*4.59 (.992)*	*-.17*
Deep			
Understanding new material for myself	4.91	5.35	
Developing as a person	4.20	5.30	
Seeing things in a different and more meaningful way	4.24	5.03	
Mean (SD)	*4.45 (.821)*	*5.23 (.851)*	*.95*

The effect-size for surface factor difference is -.17, while it is .95 for the deep factor. Teachers, thus, had a slightly weaker, though not very different, level of agreement toward the surface view of learning as students. However, teachers had a significantly greater agreement with the deep view of learning than the students. What is clear here is that teachers agreed more strongly with the deep view than the surface view of learning, yet students did not have a similar pattern. Perhaps, this could be described as students and teachers talking past each other in terms of their conceptions of learning. It is possible that the impact of the high stakes external examinations system was preventing the transfer of a deep conception of learning from teachers to students. Nevertheless, it is highly likely that students' views of learning reflect the activities that teachers provide for the students and the values that teachers express (e.g., getting good grades). Interview data reported later in this study shed light on this dilemma.

Conceptions About Curriculum

When questioned about the goals of their subject (Table 7), teachers offered responses more or less equally divided among the technological, academic, and humanistic conceptions identified by Cheung (2000). The most common goals were the

technological conception of assisting students to pass exams or maximise their

assessment results (27%), the humanistic conception of improving student life chances

or empowering them for future life (25%), and the academic conception of developing

cognitive, or subject related skills or knowledge (23%).

Table 7. Goals of Teachers for Their Subjects

Goals	Total
Technological	
Exam passing / Assessment result maximisation	14
Strive for the next level	3
Total	*17*
Academic	
Cognitive or Subject Skills developed, Understanding or Knowledge increased	12
Enjoyment of subject	4
Total	*16*
Humanistic	
Life empowerment /Increase life chances / Fulfil personal potential / Real life preparation post-school	13
Boost Confidence/sense of success/ improve motivation	3
Grow up/ Personal maturation/ increase responsibility	3
Total	*19*

The teachers in the questionnaire identified the most important learning skills in

their subject. Only six skills were selected by over half of the teachers; examination

preparation (71), analysing and learning from assessment mistakes (62), revision

procedures (62), time management (51), organisation techniques (50), and note taking

(48). This showed the overwhelming importance to teachers of ensuring that students

knew how to pass the high stakes end-of-year examinations. This emphasis seemed

strongly connected to a technological approach to curriculum.

Thus it would appear that this group of teachers had three major conceptions of

curriculum; that is technological, humanistic, and academic. It could be argued that the

humanistic and technological conceptions are not truly compatible, yet teachers

appeared able to reconcile or tolerate these differences.

Conceptions About Teaching

Teachers in the interview were asked about the various teaching approaches they used. Their responses were categorised by four themes, specifically, examination focus, social climate, methodologies, and curriculum focus (Table 8). The vast majority of responses focused on various methodologies used in the classroom, of which just under half could be associated with a transmission approach to teaching and technological approach to curriculum, (i.e., practical work and exercises, lectures and notes, model responses, students working individually). The other methods mentioned, (e.g., group or pair work, making work relevant to life or news outside the classroom, and discussion or conversation), were more classically associated with apprenticeship, nurturing, and cognitive development models of instruction. The technological, examination focus approaches (e.g., teaching examination taking techniques) when combined with the transmission approaches to teaching constituted about half of all comments about teaching (48%). Thus, teachers exhibited two major approaches to teaching; that is, a technological transmission as preparation for examinations and a nurturing, academic approach to ensure cognitive and human development.

Table 8. Teaching Approaches

Approaches	Total
Examination Focus	
Examination techniques and awareness	6
Frequent testing and assessment	4
Highly structured sequences and time frames	3
Total	*13*
Social Climate	
Provide praise and avoid put-downs	3
Firm control of student behaviour	1
Use humour	1
Total	*5*
Methods	
Practical work and exercises	9
Lectures and notes	7
Variety and diversity of activities	7
Use group or pair work	5
Make relevant to life or news outside school	4
Discussion and conversation	4
Exemplify teacher or student model responses	3
Present information in more than one way (e.g. games, AV)	3
Students work individually	2
Walk around and explain individually or to groups	2
Adapt approach to students or topic	1
Give extra teaching or tutorials	1
Total	*48*
Curriculum Focus	
Aim to develop higher order cognitive skills	3
Integrate curriculum	1
Spiral curriculum (i.e. basics, new, revision)	1
Total	*5*

Because the teachers were surveyed earlier in the year before the interviews, it was possible to ask if they had made any changes in their approaches to teaching. The type of change teachers may have made to their approach to teaching may reveal more about the conceptions of teaching they have. Teachers identified about 30 kinds of change to their teaching approaches in recent times (Table 9), with slightly more categorised as deep, student-oriented (60%) rather than teacher-controlled or examination-preparation oriented (40%). A wide variety of changes towards more constructivist or deep learning oriented practices were identified (e.g., more discussion, group work, teacher-student interaction, less copying). Simultaneously, these changes

toward student-oriented learning approaches were balanced by changes toward more

teacher-controlled transmission in preparation for success at upcoming external

examinations (e.g., reduced flexibility in content, greater control of class behaviour, less

class talk,).

Table 9. Recent Changes in Teaching Approach

Changes	Total
Teacher, Examination Oriented	
Qualification assessments make teaching less student centred or reduces topicality of teaching	5
Reduced expectations of students who are less able	2
More rigorous behavioural control or discipline	2
Less and shorter periods of talking within class	2
Extension work reserved only for self-motivated students	1
Total	*12*
Deep, Student Oriented	
More group or pair work	3
Changes to structure of lessons (e.g., breaks for writing, talk, etc)	2
Less copying of notes from board	2
More class discussion, interaction, question and answering	2
More work with ITC and AV	2
Changes to sequence of units (e.g., starting with harder material)	1
Fewer handout notes given (i.e., Students do more note making)	1
More connection of work to real life	1
More emphasis on curriculum than examinations	1
More help given to student self management of own work	1
More individual teacher-student interaction	1
Nature of notes changed (e.g., more diagrams, wacky headings, interaction)	1
Total	*18*

From the interview data, 27 of the 35 different overall descriptions of the nature

of teaching were classified as cognitive developmental or nurturing (Table 10).

Teachers largely believed the role of teaching was to develop the intellectual skills,

concepts, or knowledge of their students and to help them grow or mature as

individuals. These were student-oriented conceptions that seemed largely antithetical to

any technological emphases on examination preparation.

Table 10. Definitions of Teaching

Approaches	Total
Cognitive Developmental	
Intellectual skill, concepts, knowledge improvement	9
Love of Subject	3
Total	*12*
Nurturing	
Personal growth and development	7
Have fun / personal enjoyment	5
Make social or emotional contribution to students	4
Socio-economic life chance improvement	2
Exercise creativity	2
Total	*15*
Other	
Exercise autonomy	1
Personal survival, job orientation	2
Total	*8*

The attributes of a good teacher fell equally into two main categories (Table 11). These were understood as having to do with teachers' attitude or personality and their professionalism. According to this group of teachers, the primary personal characteristics and attitudes of a good teacher included being flexible, adaptive people who patiently forgive the troubling behaviour of young people. In addition, good teachers need professional competence in a subject and its assessment and to be able to deliver innovative, creative, well-structured, organised, and prepared lessons, based not just on their expertise but also on a love of their subject.

Table 11. Attributes of Good Teachers

Attributes	Total
Attitudes, Personality	
Patient, forgiving, empathy	6
Relaxed, flexible, adaptive	6
Approachable, personable, friendly	4
Sense of humour, fun	4
Like for or interest in students	3
Firm	2
Lively, enthusiastic, spark	2
Open attitude towards learning	2
Respectful of students	2
Commitment	1
Total	*35*
Professional Behaviour	
Expertise in field including qualification assessment	9
Flexible, innovative, creative content or classes	5
Love of subject or teaching	4
Adapting content to student interest or ability	3
Preparation, preparedness	3
Structured or organised lessons or classes	3
Ability to teach or help students to think	2
Coach or facilitate student learning	2
Listening to or interaction with students	2
Deep understanding of life and society	1
Good control of students	1
Putting curriculum, subject, or values ahead of qualifications	1
Work hard	1
Total	*37*

Thus, throughout the conversations on the nature of teaching and curriculum, there developed a two-way picture of teaching. One side emphasised humanistic and academic curriculum conceptions that seemed connected to developmental and nurturing teaching perspectives. The other side of the picture focused around technological conceptions of curriculum and transmission perspectives of teaching with the purpose of increasing students' qualification assessment results. At this point, the suspicion that the high-stakes assessment system was impacting on teachers' conceptions became more concrete. Despite an emphasis on humanistic and academic approaches to curriculum and teaching, teachers were resorting to examination swotting technologies.

Conceptions About Efficacy

Teachers were asked to identify obstacles to their achieving their curriculum or teaching goals. Four major categories of factors were identified by teachers as obstacles or hindrances to the achievement of their goals (Table 14). These were general societal background (21%), poor student behaviour or choices (53%), school structures (23%), and teacher controlled (4%) factors. The most commonly identified obstacle was poor student motivation and negative attitude towards, or lack of interest in learning and achievement. What is astonishing in this data is that so few responses (two to be exact) referred to causes within the control of the individual teacher (i.e., inadequacy or newness of the teacher or inappropriate teaching techniques). External factors underpinned the attributions of 96% of the causes offered by teachers for the failure to achieve the teachers' own goals. This indicated a high awareness of the external teacher efficacy factor was prevalent among these Year 11 teachers.

Assessment Practices

From the 82 questionnaires, a frequency ranking for the use of assessment practices in their teaching was obtained. Choices included common assessment tasks, in-class drama presentations, in-class practicals, in-class revision questions, in-class spoken presentations, in-class written assignments, in-class written tests, and long 2-3 hour written exams. A weighted total score for frequency of use was calculated by multiplying first choices by three, second choices by two, third choices by 1 and summing (Table 12).

Table 12. Frequency of Assessment Task Usage

	Frequency			Total
Assessment Tasks	First	Second	Third	Weighted
In-class written test	37	20	12	163
Common assessment task	10	26	12	94
In-class revision questions	13	16	13	84
In-class written assignment	8	13	6	56
Long 2-3 hour written exam	2	9	19	43
In-class practical	1	10	8	31
In-class spoken presentation	1	2	1	8
Other:	1	0	2	5
In-class drama presentation	1	0	0	3

The most common assessment technique used by these 81 teachers in Year 11 classes was the in-class written test selected by 69 out of 81 teachers as one of their three most common assessment tasks with a weighted score of 163. The other frequently selected assessment techniques were associated with more formal assessment, (i.e., common assessment tasks (94), in-class revision questions (84), in-class written assignments (56)). The least preferred methods of assessment were in-class practicals (31), in-class drama presentations (3), and in-class spoken presentations (8), all of which are associated with more informal assessments of learning. It would appear that this group of teachers preferred traditional examination-style formal assessments. This may be a function of their strong intention to equip students to achieve the highest possible mark in the end-of-year examinations and so implement assessment tasks that function as practice for the high stakes assessments.

When interviewed, all 18 teachers agreed that the School Certificate examination could be passed through an application of such surface strategies as rote learning, drill, or memorisation (Table 13). However, for most teachers this approach would bring about only a bare or 'C' pass in the 50s. All teachers agreed that good grades ('A' or 80% plus) at School Certificate level depended on deep understanding of the material. As well, success beyond the Year 11 School Certificate (e.g., Sixth Form Certificate or

Year 13 Bursary examinations) depended on a deep understanding of the subject, not just a rote learned surface apprehension of facts, details, or information.

Table 13. Deep and Surface Requirements in Examinations

Approaches	Total
Surface	
Surface rote learning, drill get 'C' pass	18
Rote learning can get a higher grade or 'B' pass	3
Rote learning is necessary for weaker students	2
Parts of the SC examination can be rote learned	1
Total	*24*
Deep	
Good grades depend on deep understanding @ Year 11, SFC, and Bursary,	18
Success depends on literacy, comprehension, or general knowledge	5
Rote learning is against the goals of the curriculum	2
Too much drill, rote learning creates boredom	1
Total	*26*

Conclusion

The secondary school teachers in Study 1 had a generally deep view of learning, usually focused on academic, cognitive development. They also had strong emphases on humanistic and nurturing approaches to developing the life chances, personal potential, and personal growth of students. At the same time, they also emphasised transmission approaches to teaching and technological approaches to preparation for success at the high-stakes qualification examinations or assessments. Further, teachers believed that external factors such as student attitude and motivation, societal ills, or school organisation were the main obstacles to student success. The high stakes assessments had a powerful effect on teachers' goals, practices, and conceptions about learning, teaching, efficacy, and curriculum. Thus, they exhibited largely a teacher-oriented view of schooling that seemed to excuse teachers from responsibility for students' lack of learning.

Table 14. Teacher Conceptions About Obstacles to Achieving Teaching Goals

Obstacles	Total

Societal Background	
General social ills (e.g., poverty, benefit cuts)	4
Lack of future employment opportunities	1
Poor family modelling or support	6
Total	*11*
Student Choices	
Poor motivation, attitudes, immaturity, lack of interest	11
Disruptive behaviour/ lack of self-discipline	4
Negative perception of subject or teacher	3
Lack of prior knowledge or skills	3
Substance abuse	2
Lack of effort (e.g., not doing homework)	2
Absenteeism	1
Too much paid work	1
Lack of ability	1
Total	*28*
School Control	
Class sizes too large	3
Insufficient time with students	2
Over assessing	2
Lack of school resources	2
Crowded curriculum	1
Examination structure or system	1
School timetabling	1
Total	*12*
Teacher Control	
Inadequacy or newness of teacher	1
Inappropriate teaching techniques	1
Total	*2*

This study also successfully demonstrated the usefulness of the six-point agreement response scale to identify teachers' conceptions. Teachers' attitude scale responses seemed to be similar to their conceptions as expressed in interviews. The six-point response scale successfully provided adequate variance in teachers' responses so that distinctions between strength of conceptions could be identified.

However, this study did not provide insight into the structure of teachers' conceptions of assessment; rather, not unexpectedly, it pointed out the power of the assessment system to influence teachers' practices and conceptions about learning, curriculum, efficacy, and teaching. Further research into the nature of teachers' conceptions of assessment was needed.

Study 2: In-Service Primary Teachers' Conceptions of Assessment

It was obvious from a study of secondary teachers who work in an environment of high-stakes external examinations that the assessment system had a strong impact on teachers' conceptions about teaching, learning, and efficacy. Furthermore, the views of primary school teachers, where the impact of high stakes qualifications assessment is largely irrelevant, seemed necessary to complement the picture developed in Study 1 where teachers exhibited deep views of learning with technological and transmission approaches to teaching and curriculum perhaps as a consequence of the need to ensure student success at those examinations. Thus, this study sought a more thorough understanding of how teachers conceive of assessment and how those conceptions relate to the conceptions discussed in the literature review.

This study was designed to collect open-ended, free-response data from practising primary school teachers about how they understood assessment and then to assess the adequacy of the four conceptions noted from the literature to classify the responses. The other goal was to elicit conceptions that may not have been discovered in the literature review.

Instrument

Data were obtained from teachers who had just administered a standardised paper-and-pencil student assessment of either reading or writing. This process permitted teachers an opportunity to express their opinion in response to a certain well-understood style of assessment. The reading tests were a mixture of constructed response and multiple-choice questions requiring a combination of surface and deep

cognition. The writing tests involved students in completing dictated and edited spelling tasks and an extended piece of writing that focused on one main purpose of writing. All assessment tasks were designed for classroom use with an improvement conception of assessment in mind, as part of a cycle of collecting information in order to diagnose student learning needs, describe student strengths, and implement an instructional programme that would address learning needs. The assessment tasks were being calibrated as part of the preparation of materials for a CD-ROM item bank for teacher controlled, computer-assisted test creation, analysis, and reporting. Further details of the characteristics of the reading assessment tasks can be found in Meagher-Lundberg and Brown (2001), and details of the writing assessments can be found in Glasswell, Parr, and Aikman (2001). Teachers were asked to explain to what extent the assessment they had administered related to their own view of assessment and to provide up to a paragraph explaining in what ways their views differed or were identical to the type of assessment they had just administered.

Participants

To obtain data, a large-scale survey of in-service teachers' conceptions about assessment was conducted. Around 20,000 assessments were sent out to a nationally representative sample of approximately 500 schools, with half the schools administering reading assessments and the other half administering writing assessments. On the basis of about one teacher per 25 students, the potential population of teachers participating was estimated to be 800. Just over half of the estimated population of teachers ($N = 465$) replied to the questionnaire (response rate = 58%), with slightly fewer responding to the reading assessments ($n = 218$) than the writing assessments ($n = 247$).

Results

Just over 500 comments were relevant to the topic of assessment and were classified into the four main assessment purpose conceptions (i.e., improvement, irrelevance, school accountability, and student accountability) (Figure 1). In addition, statements could be assigned to two learning (i.e., deep and surface) conceptions and two teaching (i.e., examination and nurturing) conceptions. Just five statements could not be classified into any of the purpose conceptions and so were deleted from this analysis. Overwhelmingly, teachers' comments were of the improvement type (62%), with irrelevance comments being the next common type (19%).

Figure 1. Primary Teachers' Conceptions of Assessments Comments by Category
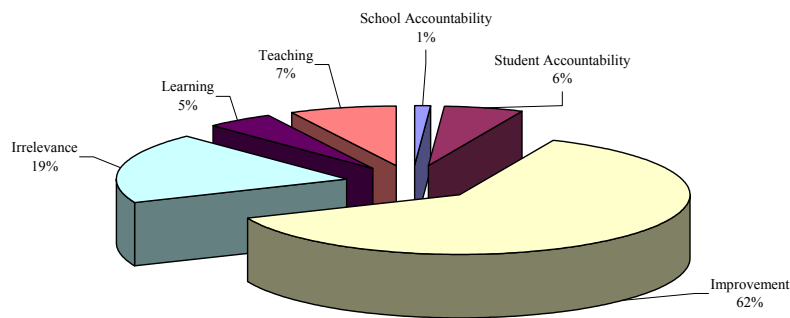
Table 15. In-Service Teachers Conceptions of Assessment Organised by Four Major Conceptions

Conceptions and Statement	Reading	Writing	Total
School Accountability			
Assessments confirm/moderate teacher judgements	1	1	2
Assessments evaluate how teachers, schools, and systems are performing	1	1	2
Assessment provides useful external reference point	0	1	1
Assessment results measure teacher or school effectiveness	0	1	1
Sub-Total	*2*	*4*	*6*
Student Accountability			
Assessment describes components of performance using standards, levels, or criteria	0	20	20
Assessment compares students one to the other, creating benchmarks for student achievement	1	8	9
Sub-Total	*1*	*28*	*29*
Improvement Describe			
Assessment describes or identifies or establishes the abilities of students or what they know and can do or what they have learned to know or do or what they have remembered or shows what level they are at	18	16	34
Assessment identifies student strengths and weaknesses, errors, gaps or needs; provides diagnostic or formative information	13	6	19
Assessment, informed by criteria and results, determines how much students learned of teaching	1	4	5
Assessment is a process to collect information for reporting, research, improvement, accountability, etc.	1	1	2
Assessments allow valid inferences about student ability or knowledge	1	0	1
Sub-Total	*34*	*27*	*61*
Improvement of Student Learning			
Feedback from assessment is communicated to students	0	2	2
Sub-Total	*0*	*2*	*2*

Conceptions and Statement	Reading	Writing	Total
Improvement Quality of Assessment			
Assessments use a variety of materials, methods, question and response formats	53	35	88
Formative assessment is carried out using multiple techniques	13	10	23
Assessment is multi-faceted having formal and informal techniques; involving self, peer, and group assessment techniques	7	15	22
Assessments replicate real world events, relevant, practical, realistic	14	5	19
Assessment must be manageable	10	8	18
Assessment is standardised	2	3	5
Assessment is objective	1	1	2
Scoring of assessments determines whether the assessment focuses on deep or surface learning	0	1	1
Assessment has to be systematically carried out	1	0	1
Assessment is fair and unbiased	1	0	1
Consistency or reliability is important in assessment	1	0	1
Good assessments take time and effort to create	1	0	1
Continuous assessment is better than one off	1	0	1
Sub-Total	*105*	*78*	*183*
Improvement of Teaching			
Assessment is integrated with teaching and curriculum; testing and teaching use similar activities	9	32	41
Assessment aids planning or determines what next teaching or further learning is required by providing feedback to teachers	8	14	22
Assessment is a basis for grouping students for differential instruction	4	1	5
Assessments define and exemplify learning and performance	2	0	2
Assessment usually aims at improvement	1	0	1
Formative assessment is integrated with constructivism	1	0	1
Sub-Total	*25*	*47*	*72*

Conceptions and Statement	Reading	Writing	Total
Irrelevance Bad for Students			
Assessment provides a snapshot of student ability and is unfair because it is not a full picture of student ability	13	17	30
Assessment must be fair to children in terms of preparation, teaching, etc.	6	1	7
Tests provide information out of context	2	3	5
Assessment is not connected to students' real ability, just their test taking ability	1	1	2
Sub-Total	*22*	*22*	*44*
Irrelevance Assessment is Inaccurate			
All assessment is flawed	1	1	2
Assessment results must be used carefully and cautiously	2	0	2
Assessment results are fuzzy and approximate, especially where marker judgement is involved	0	1	1
Assessments are too reliant on reading and writing skills	1	0	1
Observational tools are better than tests	1	0	1
Sub-Total	*5*	*2*	*7*
Irrelevance to Teaching			
Assessment is done best in the head of the teacher using professional observation and judgement against learning outcomes or criteria or standards as part of regular classroom life	7	14	21
Assessment is not connected to real learning or classroom activities	10	2	12
Assessments are not necessarily aligned to teaching or curriculum	11	0	11
Assessment has negative consequences on teaching	2	0	2
Assessment results are filed and ignored	0	1	1
Assessments are not needed to guide teaching; curriculum and experience is enough	1	0	1
Sub-Total	*31*	*17*	*48*
Learning Deep			
Assessment measures higher order thinking skills	10	0	10
Sub-Total	*10*	*0*	*10*

Conceptions and Statement	Reading	Writing	Total
Learning Surface			
Assessment measures lower order skills	4	11	15
Sub-Total	*4*	*11*	*15*
Teaching for Examination			
On-going assessments can act as practice for higher levels of schooling	0	2	2
Sub-Total	*0*	*2*	*2*
Teaching Nurturing			
Assessment must be child-centred; friendly, not too difficult, manageable, build confidence, remove anxiety, informs of criteria and results	14	9	23
Assessment engages student interest, enjoyment	9	0	9
Assessments are individual activities, not group or pair work	1	1	2
Teacher scaffolding is good in assessment	2	0	2
Sub-Total	*26*	*10*	*36*

Table 15 lists the 13 assessment, learning, and teaching conceptions detected in the open-ended comments by the teachers. Each conception is exemplified by groups of statements that constitute various sub-categories within the conception. For example, the improvement conception has four sub-categories (i.e., assessment describes student learning, the quality of assessment tasks, student learning, and teaching) while the irrelevance conception has three sub-categories (i.e., bad for students, inaccurate, and teaching). The various conceptions, sub-categories, and statement groups are listed by whether the teacher had administered either a writing or reading assessment. Note that only 17 statement groups had 10 or more respondents, and that just 10 statement groups had 20 or more respondents. The ten most frequent statement groups represented 324 or 62% of all interpretable comments, while the next seven frequent gave 104 or 20% of all comments. Thus, nearly three-quarters of all responses were captured by those 17 statement groups and are discussed in this section.

Three of the four most common statement groups with frequencies of 30 or more were related to the improvement conception and one was related to the irrelevance conception of assessment. The improvement statement groups focused on the quality aspects of assessment tasks, their descriptive power, and their ability to improve the quality of teaching. Specifically, these most frequent statements showed that teachers believed that (a) assessment must use a variety of materials, and methods, including diverse question and response formats (88), (b) assessment is integrated with teaching and curriculum so that similar activities are used in assessment and instruction (41), and (c) assessment describes, identifies, or establishes the abilities of students, what they know and can do, what they have learned to know or do, what they have remembered, what level they are at (34). The

irrelevance conceptions focused on how assessment is bad for students, specifically assessment provides only a snapshot of student ability and is, thus, unfair because the result is not a full picture of student ability (30). This nexus of conceptions, if taken together, revealed a perspective that viewed assessment as multiple means of obtaining a single-snapshot description, closely linked to teaching of student achievement.

The next six most common comments, with frequencies between 20 and 29, extended the improvement conception of assessment by focusing on the quality dimension of assessment tasks. Diversity of assessment technique (23), including the use of informal, peer, self, and group assessments (22), was emphasised. The nurturing teaching perspective was seen in the frequently expressed statement that assessment must be student-centred; student-friendly, not too difficult, manageable for students, build their confidence, and remove their anxiety (23). The role assessment plays in improving the quality of teaching was expressed in comments that focused on assessment as an aid to planning or determining the next teaching or further learning steps (22). This dimension was supported by a continuing strong emphasis on assessment that provides describes the components of student performance compared to standards, levels, or criteria (20). The irrelevance view was extended by comments that assessment is best done in the head of the teacher using professional observation and judgement against learning outcomes or criteria or standards as part of regular classroom life (21).

The last set of comments (i.e., seven comments made between 10 and 19 times) continued the improvement themes of assessment as a source of descriptive information about student strengths, weaknesses, errors, gaps, or needs (19), that assessment uses realistic, relevant, and practical activities (19), that its

implementation is manageable (18), and that it provides information about students'

higher order thinking (10). In contrast, there was a frequently expressed view that

assessment was irrelevant to teaching because it was perceived as not connected to

real learning (12) or classroom activities or not aligned to teaching or curriculum

(11).

It is worth noting that about seven percent of comments could be attributed to

either the school or student accountability conceptions. Thus, though they were not

frequently expressed overall, it does appear that teachers were aware of the

accountability conceptions. However, in the context in which the data was being

collected, those conceptions were not paramount. Further, it should be noted the

some teachers associated assessment with learning conceptions, though 3:2 in favour

of a surface conception.

This sample of teachers viewed assessment in largely improvement terms

(e.g., obtaining diagnostic evidence about student achievement in order to shape

teaching and learning activities). This was tempered by the view that assessment was

irrelevant to teaching, learning, and students, and the conception that assessment was

part of school accountability.

Conclusion

Study 2 identified two major conceptions held by primary school teachers

about assessment (i.e., improvement and irrelevance). These conceptions were multi-

faceted with three irrelevance and four improvement themes identified in the

literature being seen in their comments. For example, the improvement conception

was based on a diagnostic description of student performance using high quality

tasks, integration of teaching and assessment, and the use of assessment in planning

instruction. The irrelevance conception identified assessment as irrelevant to teaching, as something bad for students, and as inaccurate. This irrelevance conception may have been closely linked to a student-centred teaching approach. It was also evident that asking teachers to discuss their conceptions of assessment triggered comments related to their conceptions of teaching and learning.

Study 1 showed that secondary school teachers' conceptions of teaching and learning were influenced by the high stakes student accountability examinations. In this study, information about school-wide assessment systems was not sought and so no comparison could be made. The open-ended comments supported on the whole the range of conceptions developed from the literature and suggested that teachers' conceptions were complex and multi-structured rather than simple and dichotomous.

A serious limitation in both Studies 1 and 2 is that the nature of the association or linkage of ideas described so far has been imposed or constructed by the researcher. Teachers did not provide explicit information about how strongly they held each of their conceptions of assessment or how they would relate one conception to another; rather they simply indicated that they had one or more conceptions of assessment. A more sophisticated approach to data analysis, and consequentially data collection, was needed to establish in detail not only what conceptions teachers had about assessment, how strongly they held the varying conceptions, and how those conception related to the various assessment conceptions they held. Hence, it was decided to conduct a series of studies that would help expose the associations between and strengths of teachers' conceptions of assessment. First, an instrument to elicit teachers' conceptions of assessment had to be developed. Second, pre-existing instruments to measure teachers' conceptions of curriculum, teaching, and teacher efficacy had to be found.   Third, data had to be collected from much larger samples

to assist with analysis and generalisability. Fourth, a more sophisticated way of testing the proposed models of how teachers' conceptions interrelate had to be used. The technique most suited for the type of data being collected is measurement modelling based on structural equation model techniques.

The next chapter reports three studies that developed a new instrument to measure teachers' conceptions of assessment. The subsequent chapter then reports on the instruments used to explore how teachers' conceptions of assessment relate to other constructs and the results of that research.

CHAPTER III. MEASURING TEACHERS' CONCEPTIONS OF ASSESSMENT

Although studies 1 and 2 were able to paint a picture of how assessment influences teachers' learning and teaching conceptions and how assessment was seen as tied to the improvement of learning, those studies were not able to flesh out the structure of teachers' conceptions of assessment. To be able to answer the questions of what conceptions of assessment teachers hold, how strongly they might agree with a conception, and how the various conceptions related to each other, it was necessary to develop an instrument that could manageably, validly, and reliably provide an estimate of teachers' conceptions. An instrument that provides reliable scores to a theoretically valid mapping of a construct, such as teachers' conceptions of assessment, would also be valuable for future research and professional development activities. However, it was not possible to locate an instrument to measure teachers' conceptions of assessment in the literature.

Thus, standard processes for the development of an attitude scale (Gable & Wolf, 1993) were implemented and are documented in this chapter. These processes involved turning conceptions of assessment into statements to which participants could indicate their degree of agreement or disagreement. Then, exploratory factor analyses and confirmatory (or rather restricted) factor analyses were used to develop and test a model of how teachers' conceptions of assessment was structured. The process of developing an instrument depends on having and testing a model of how the domain of interest is structured. The danger in using sophisticated analytic techniques, that allow post-data collection manipulation, is that statistically significant results may be due to chance factors within the population being studied. Thus, it is advisable to use multiple samples to test each new version of an instrument

95

being developed (MacCallum, 1995). Consequently, this chapter reports three studies conducted with different education-profession populations. An instrument is also able to estimate the strength of attitude teachers have towards each construct as measured by teachers' responses to the statements. These scale scores are useful in testing whether there are differences in conceptions among different populations and for linking conceptions of assessment to other relevant constructs.

Thus, this chapter reports three studies into the development of an instrument to measure teachers' conceptions of assessment. Study 3 used pre-service primary and secondary postgraduate teacher trainees, while study 4 used undergraduate education students, and study 5 used in-service primary teachers. These studies used self-report Likert-type response questionnaires. In each study a Likert-type response scale used in Study 1 was adopted (i.e., a six-point response scale involving 4 positive and 2 negative agreement responses), with identical scoring (i.e., items were scored 1 to 6 with extremes representing respectively strong disagreement and strong agreement). All items were written in a positive frame so that reverse scoring was not needed. The questionnaire was trialled three times with three different education-profession samples, and analysed to produce an inventory of statements that exemplified various conceptions of assessment.

The confirmatory factor analytic technique used to determine the nature of the structure of teachers' conceptions of assessment was structural equation modelling (SEM). SEM allows tighter specification of multiple hierarchies or paths between factors by utilising the factor patterns, correlation patterns, covariance patterns, and residual values within a data matrix (Hoyle, 1995). Specification of a model includes identifying observed variables that load onto latent first-order factor, and the relationship of the first-order factors to second or higher-order latent factors. It is

critical to remember that the development of an SEM model is dependent on theory, in other words SEM is a sophisticated correlational technique that ought to be used only in the context of a proposed meaningful set of relationships (Maruyama, 1998). Structural equation modelling generates two types of model (i.e., measurement and structural) wherein measurement models explain the relationships among structures while structural models predict relationship between factors (Hoyle, 1995).

Choosing goodness of fit indicators of these models is still contentious, but there is general agreement that the more effective measures (i.e., least affected by sample size) are when the Tucker-Lewis Index (*TLI*), and Adjusted Goodness of Fit Index (*AGFI*), are greater than .90 and the absolute fit of the model, Root Mean Square Error of Approximation (*RMSEA*) is below .08 (Hoyle, 1995). Sample size is also critical as the number of parameters increases (Browne & Cudeck, 1989; 1993), with numbers greater than 500 recommended for most cases (Chou & Bentler, 1995). It is also desirable that the factor loadings are much greater than zero (indicated by the loadings being at twice their standard errors), and that no modification index points to a factor loading as being inappropriately fixed to zero. It is worth noting that, although modification indices are powerful in increasing fit of a model to the data when used in accordance with a meaningful theory, such modified models need to be tested on new samples to ensure that the models are not dependent on chance characteristics of the data set being used (MacCallum, 1995). SEM analysis in this thesis was conducted with AMOS (Arbuckle, 1999).

Study 3: Teacher Trainees' Conceptions of Assessment (CoA-I)

Statements that represented major ideas identified in the review and from the teacher comments in Study 2 were generated and incorporated into a questionnaire form. It should be noted that the instrument had to be manageable in length and so a restricted number of statements (115) was essential to ensure cooperation of voluntary, unrewarded participants. The implication is that multiple studies were needed to trial items for all the various conceptions identified in the literature. It was found that most participants completed the full questionnaire (approximately 100 rated statements and demographic questions) in 20 to 30 minutes. Two forms were used with reversed order of statements to take account of any participant fatigue.

The first trial of the Conceptions of Assessment (CoA-I) inventory made up of 115 statements was completed by 84 primary and secondary pre-service Diploma of Teaching trainees. Because most items (101/115) had less than 5% missing at random data ($M = 3.28\%$), missing value analysis, using the SPSS EM procedure, was conducted to provide 84 complete response sets (SPSS, 2000). As the sample size (particularly in relation to the number of items) was not sufficiently high, principal axis rather than maximum likelihood factor analysis with oblimin rotation was conducted with items grouped according to three of the major meta-concepts (i.e., accountability, improvement, and irrelevance). For a factor to be retained it had to have at least 3 statements loading >.30 and for a statement to be retained it had to have a clear logical connection with other statements loading on the factor and have a loading of >.30. Factor reduction led to ten factors and 65 statements (Table 16).

Table 16. Conceptions of Assessment Factors, Statements, and Loadings

Factors and Statements	Loading
High Stakes Testing is Bad for Teaching	
Teachers pay attention to assessment only when stakes are high	.73
Teachers are over-assessing	.68
Assessment forces teachers to teach in a way against their beliefs	.57
High stakes assessment has negative impact on teaching	.57
Classroom assessment is practice for high-stakes testing	.42
Assessment Serves External Stakeholders	
Assessment selects students for future education or employment opportunities	.78
Assessment in secondary schools is dominated by external or public examinations	.64
Assessment results should be communicated to students and their parents	.62
Teachers are evaluated by assessment	.61
Assessment shapes the activities that students undertake	.51
Assessment provides information on how well schools are doing	.48
Provides Useful Information About the Quality of Teaching	
Assessment moderates teacher judgement about student performance	.73
Assessment confirms or moderates teacher judgement	.71
Assessment results measure teacher effectiveness	.60
Assessment provides useful external reference points of what student achievement should look like	.47
Assessment gives information on how well teachers teach	.37
Assessment Improves Teaching	
Assessment models useful teaching strategies	.75
Assessment improves teaching	.69
Assessment is all about feedback to students about performance	.64
Assessment is an essential part of teaching and curriculum	.62
Assessment is integrated with constructivist teaching practice	.62
Assessment Diagnoses Student Ability, Knowledge, Learning, Thinking	
Assessment permits valid interpretations of what students can do	.77
Assessment identifies how students think	.69
Assessment identifies student strengths and weaknesses	.69
Assessment establishes what students have learned	.69
Assessment is a way to determine how much students have learned from teaching	.68
Answers to assessment show what goes on in the minds of students	.62
Assessment provides information about how students are performing	.59
Assessment identifies what students know or can do	.56
Assessment provides recommendations about student learning	.51
Assessment explains why students perform in a certain way	.40
Assessment defines and shows what learning should take place	.38

Factors and Statements	Loading
Assessment Improves Students' Learning	
Assessment is a positive force for improving social climate in a class	.69
Assessment is an engaging and enjoyable experience for children	.59
Assessment describes the abilities of students	.55
Assessment measures higher order thinking skills	.54
Assessment is appropriate and beneficial for children	.51
Assessment helps students improve their learning	.49
Assessment Shapes Teaching	
Assessment aims at improvement of student learning	.49
Assessment allows different students to get different instruction	.45
Assessment is a way to determine what next teaching is required	.65
Assessment changes the way teachers teach	.51
Assessment influences the way teachers think	.54
Assessment is Ignored	
Teachers should ignore assessment	.83
Assessment is unfair to students	.73
Assessment has little impact on teaching	.72
Assessment has negative consequences on teaching	.70
Assessment results are filed and ignored	.66
Assessment is value-free	.51
Assessment Has Limitations	
Assessment is too reliant on reading and writing	.79
Assessment is limited because it does not show the full picture of a student's abilities	.69
Success in assessment is due to test taking ability not real ability	.62
Assessment is biased against minority group students	.53
Observation is the best type of assessment	.47
Assessment measures memorisation and rote learning	.42
Assessment is Valid	
Assessment results can be depended on	.82
Assessment results are consistent	.65
Assessment results predict future student performance	.65
Assessment results are trustworthy	.64
Assessment activity is manageable	.62
Assessment is objective	.58
Assessment concludes with an overall evaluation of the student's learning	.50

The ten factors were broken into three main areas, (i.e., accountability, improvement, and irrelevance). The two accountability factors were (a) assessment provides useful information to teachers about teaching and (b) assessment serves external stakeholders. The five improvement factors were (a) assessment improves teaching, (b) assessment diagnoses student ability, knowledge, learning, and thinking,

(c) assessment is good for students, and (d) assessment shapes teaching, and (e) assessment is valid. The three irrelevance factors were (a) assessment is ignored, (b) assessment has limitations, and (c) high stakes testing is bad for teaching. Once factors were found mean scores and estimates of reliability (coefficient alpha) for each factor were generated (Table 17, Figure 2). Multiple analysis of variance revealed no statistically significant differences between males and females, primary and secondary trainees, or between ethnic groups for any of the sub-scale scores.

Figure 2. CoA-I Factors Mean Scores
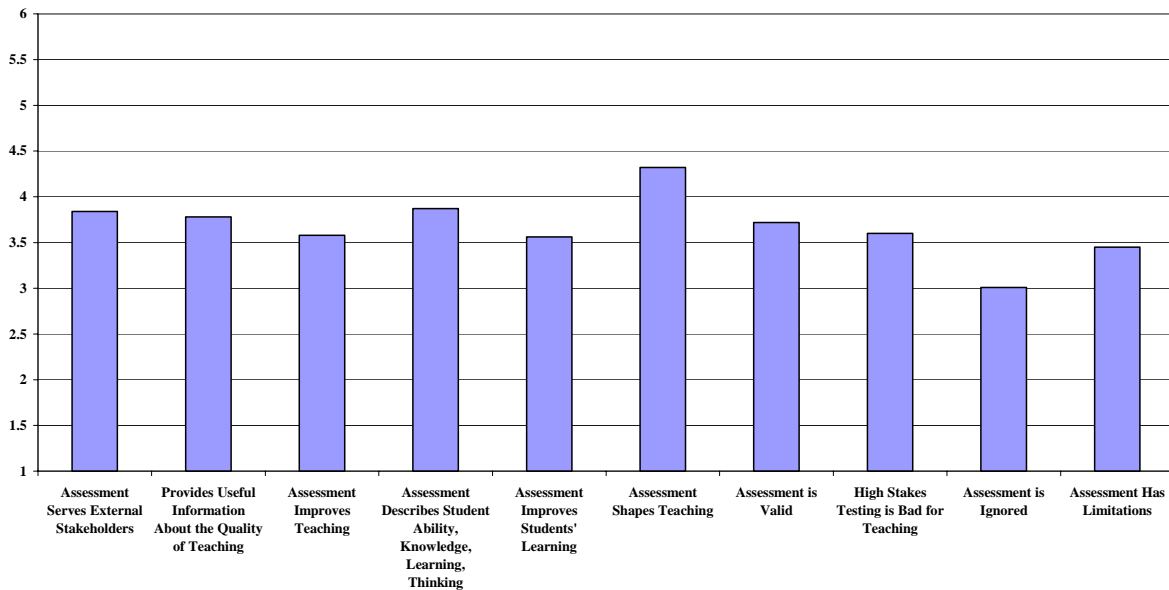
Teacher trainees' mean scores for all ten factors were close to each other within the range of 3.5 to 4.0 (i.e., close to moderately agree), except for two scales. They more than moderately agreed that assessment shapes teaching and only just slightly agreed that assessment is ignored. Thus, for this group of future educators, they agreed with the improvement (i.e., mean score for five sub-scales was 3.81) and accountability (i.e., mean score for two sub-scales was 3.81) conception factors. In contrast, the irrelevance factor received weak agreement (i.e., mean score for three

101

sub-scales was 3.35). Estimates of reliability for the sub-scales were moderate to

good (i.e., alpha range from .61-.82) with an average alpha of .73.

Table 17. CoA-I Factors Means and Reliabilities

Factor	Estimate of Reliability (α)	Mean (*SD*)
Accountability		
Assessment Serves External Stakeholders	.73	3.84 (.15)
Provides Useful Information About the Quality of Teaching	.68	3.78 (.11)
Improvement		
Assessment Improves Teaching	.76	3.58 (.31)
Assessment Describes Student Ability, Knowledge, Learning, Thinking	.82	3.87 (.11)
Assessment Improves Students' Learning	.73	3.56 (.36)
Assessment Shapes Teaching	.61	4.32 (.04)
Assessment is Valid	.81	3.72 (.24)
Irrelevance		
High Stakes Testing is Bad for Teaching	.72	3.60 (.02)
Assessment is Ignored	.75	3.01 (.10)
Assessment Has Limitations	.68	3.45 (.05)

Note that the meaning of the irrelevance items and factors needs to be

understood clearly; low scores indicate that a statement or factor is relevant (not

irrelevant) while high scores indicate that a statement or factor is irrelevant. As the

average score was greater than the mid –point of the Likert scale, this group of

teacher trainees agreed with most conceptions of assessment, though there was greater

agreement with accountability and improvement conceptions than irrelevance

conceptions.

However, the small sample size meant that it was not possible to analyse all 65

statements and 10 factors at one time and thus it was not possible to establish the

nature of the relationship between the various conceptions. Further trialling with

larger populations was required to test whether the assignment of factors to meta-

concepts was merited. Another limitation was that no practising teachers completed

the questionnaire. Practising teachers were included in Study 4.

Further, nearly half the items trialled in this study did not fit the model, and so another trial would permit testing of further focused statements around conceptions in the literature that had not been adequately detected in this study. This meant that items not found to fit the model in this study would be dropped and new items tested with these items in the next study. Study 4 involved an extra 40 statements derived from further searches of the literature. Sources for further conceptions and items included the models of teacher assessment practices documented by Gipps, et al. (1995) and the doctoral dissertation of M. Hill (2000b) and an article on secondary teacher assessment practices (McMillan, 2001b). In addition, statements that focused on peer and self-assessment, and academic enabling (e.g., positive attitude, hard effort, etc.) were added. Thus, a second version of the CoA-I was developed for Study 4 that had 105 statements.

Study 4: Students' and Practising Teachers' Conceptions of Assessment (CoA-II)

As noted earlier, each of the three studies reported in this chapter selected items that fit the model and then introduced new items based on the model to more fully test the model. Study 3 resulted in 65 statements loading on to 10 factors, each of which had been tested independently of each other. A second instrument (CoA-II) of 105 statements was assembled utilising the 65 statements from Table 17 and 40 more statements that had been written to pick up key ideas not previously included in the instrument.

Participants ($N = 188$) were 141 education students in either first year developmental psychology or second year education sociology and 47 practising primary teachers. Missing data (no more than 2-3% for each variable) were imputed with the SPSS EM missing value procedure and then corrected to a minimum of 1.0 and maximum of 6.0. Because of the low cases to variables ratio, three maximum likelihood factor analyses with oblimin rotation were conducted and subsequently tested as three measurement models (i.e., accountability, improvement, and irrelevance). Only the statements that were expected to load on each of the three major assessment conceptions (i.e., accountability, improvement, irrelevance) were analysed together. Thus, clearly a significant limitation of this study is the inability to test a model that integrates the three measurement models.

The loadings of the 46 CoA-II statements retained by the factor analysis on each first order factor are shown in Table 18. Note that 11 items in the present 46 are new in this modelling; meaning 35 statements were carried over from the Study 3. All statements had loadings on their respective first order factor greater than .40.

Table 18. CoA-II Factor Structure Study 4

First Order Factor and Statement	Loading
Accountability: Evaluate Learning Objectives	
Assessment is completing checklists	.70
Assessment is comparing student work against set criteria	.66
Assessment is assigning a grade or level to student work	.66
Assessment is systematic collection of information about student achievement	.63
Assessment is checking off progress against achievement objectives	.58
Accountability: Evaluate Schools	
Assessment provides information on how well schools are doing	.88
Accountability: Evaluate Students	
Classroom assessment is practice for high-stakes testing	.79
Assessment shapes the activities that students undertake	.73
Assessment selects students for future education or employment opportunities	.53
Accountability: Evaluate Teachers	
Assessment results measure teacher effectiveness	.66
Assessment provides teachers useful external reference points of what student achievement should look like	.61
Assessment gives teachers information on how well they teach	.57
Improvement: Improve Student Learning	
Assessment helps students improve their learning	.79
Assessment is appropriate and beneficial for children	.75
Assessment feedbacks to students their learning needs	.68
Assessment provides feedback to students about their performance	.66
Assessment is an engaging and enjoyable experience for children	.66
Assessment makes students do their best	.61
Assessment is a positive force for improving social climate in a class	.60
Improvement: Improve Teaching	
Assessment information modifies ongoing teaching of students	.67
Assessment is integrated with teaching practice	.62
Assessment changes the way teachers teach	.59
Assessment allows different students to get different instruction	.58
Assessment information is collected and used during teaching	.51
Assessment influences the way teachers think	.48

First Order Factor and Statements	Loading
Improvement: Diagnose Student Ability	
Assessment measures students' higher order thinking skills	.76
Assessment establishes what students have learned	.75
Assessment identifies student strengths and weaknesses	.72
Assessment is a way to determine how much students have learned from teaching	.69
Assessment identifies how students think	.66
Answers to assessment show what goes on in the minds of students	.56
Improvement: Quality Valid Information	
Assessment results are trustworthy	.75
Assessment results can be depended on	.59
Assessment results are consistent	.57
Assessment is objective	.49
Assessment results predict future student performance	.42
Irrelevance: Bad for Teaching	
Assessment is unfair to students	.71
Teachers are over-assessing	.67
Assessment interferes with teaching	.63
Assessment forces teachers to teach in a way against their beliefs	.62
High stakes tests are bad for teaching	.54
Assessment is biased against minority group students	.54
Teachers pay attention to assessment only when stakes are high	.47
Irrelevance: Used but Ignored	
Assessment is value-less	.73
Assessment has little impact on teaching	.49
Assessment results are filed and ignored	.44

Factors that had more than three statements with related meaning and that loaded >.30 on the factor, were tested with measurement modelling to establish degree of fit for each of three measurement models (i.e., accountability, improvement, and irrelevance). A higher-order model involving first-order factors and second-order factors for each measurement model was tested to determine if data conformed to the proposed hierarchical models. The accountability and improvement measurement models had four first order factors loading onto one second-order factor (Figures 1 and 2), while the irrelevance model (Figure 5) had two first order factors loading. This gave a total of 10 first order factors loading onto three second-order factors in three unrelated measurement models.

Table 19 shows the three measurement models with factor names, number of items in each factor, the average score per item, the loading of each factor on its appropriate second order factor, the internal reliability of each factor scale, and the average loading of statements on each factor. Six of the first order factors had between five and seven variables, three first order factors had only three variables, and one factor (i.e., Accountability: Evaluate Schools) is made up of a single variable. The quality of these models was apparent from inspection of the fit indices. The data showed that the models for Irrelevance and Improvement have reasonable fit to the data, while those for Accountability were close to reasonable fit but required further work (*RMSEA* > .08); certainly more items for *Evaluate Schools* factor are needed. Another possible source of poor fit is the possibility that the accountability conception is actually constituted of two separate but correlated conceptions (i.e., one related to schools and a second to students). Further analyses with more items and larger sample size are required to establish whether the accountability conception exists as one construct with two factors or as two separate conceptions.

Table 19. CoA-II Measurement Model Characteristics: Accountability, Improvement, and Irrelevance

Second-Order (χ^2; *df*; *RMSEA*; *TLI*) First-Order Factors	*N* of items	Characteristics			
		1st Order Factor Scale *M*	Loading on 2nd Order Factor	1st Order Factor Scale Reliability (α)	1st Order Factor Item Loading *M*
Accountability (129.83; 51; .091; .843)					
Evaluate Teaching	3	3.75	.60	.64	.61
Evaluate Schools	1	4.14	.88	na	na
Evaluate Students	5	3.88	.62	.71	.68
Evaluate Learning of Objectives	3	4.15	.77	.78	.65
Improvement (533.14; 248; .078; .763)					
Improve Teaching	6	4.10	.37	.75	.58
Improve Learning	7	3.54	.95	.86	.68
Quality Validity	5	3.20	.88	.69	.56
Describe Achievement	6	3.72	.99	.85	.69
Irrelevance (60.88; 34; .065; .908)					
Bad for Teaching	7	3.40	.52	.80	.60
Used but Ignored	3	2.28	.55	.57	.55

Loadings of items onto first order factors are high with actual loadings ranging between .42 and .88. Loadings from the first order factors to the second order factors range from .37 to .99, indicating that first-order factors provide reasonably similar contribution to the second-order factor. Reliabilities for each first order factor scale are moderate (i.e., .57 for *Used but Ignored* having only 3 items) to strong (i.e., .86 for *Improve Learning* having 7 items) with the average alpha being .74. Note that reliability for *Evaluate Schools* could not be calculated because only one item makes up this factor. The data for the three measurement models are diagrammed below; accountability (Figure 3), improvement (Figure 4), and irrelevance (Figure 5).

Figure 3. CoA-II Accountability Conception Measurement Model
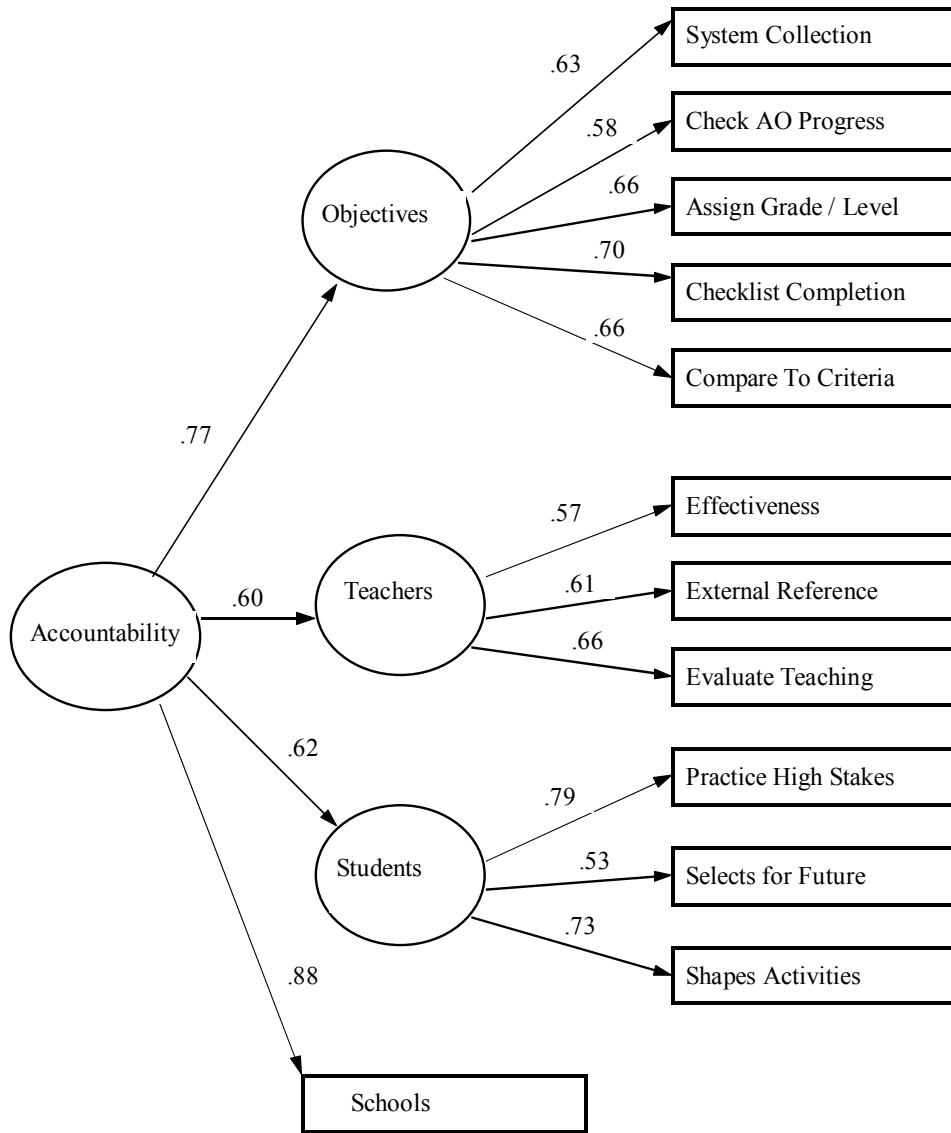
Figure 4. CoA-II Improvement Conception Measurement Model
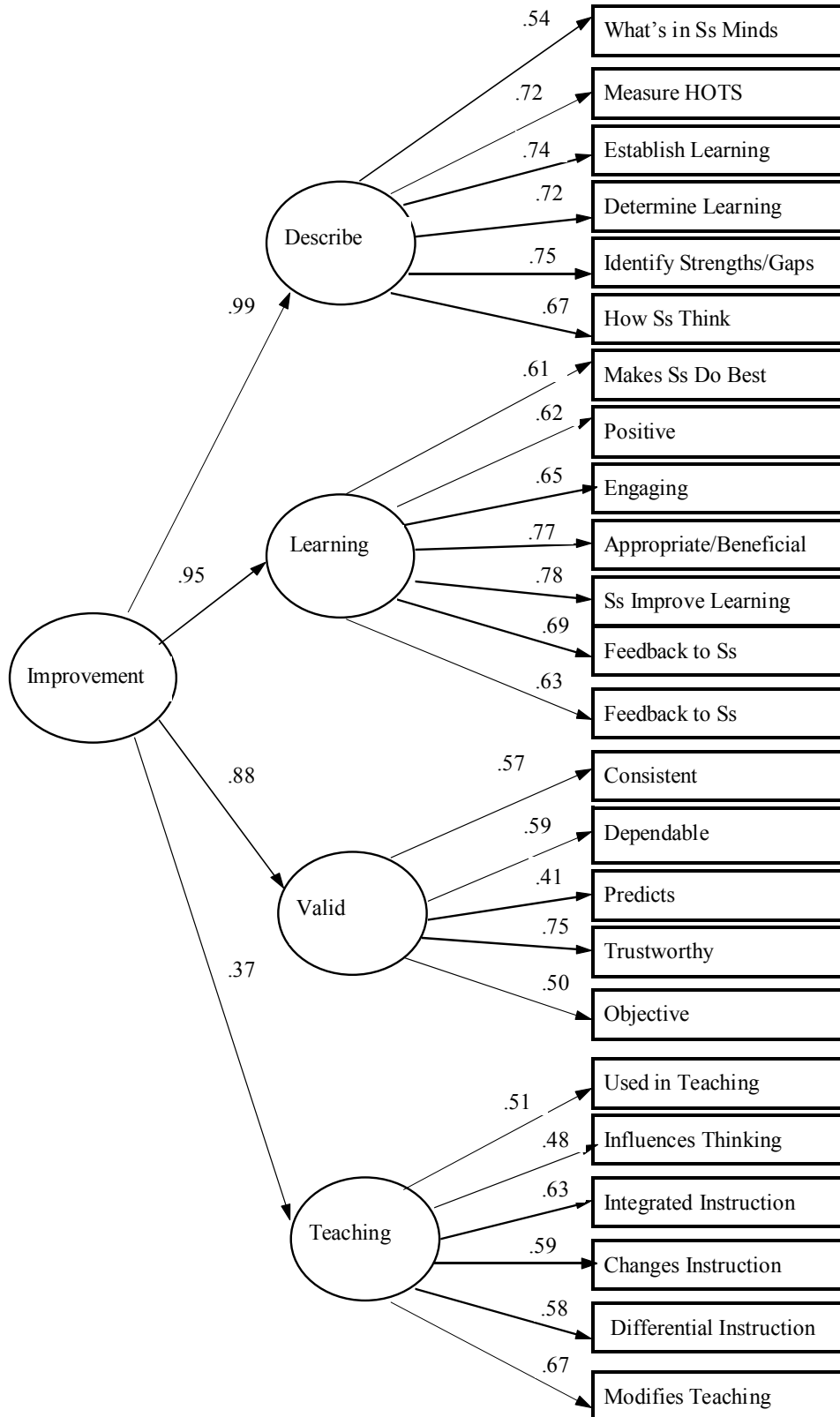
Figure 5. CoA-II Irrelevance Conception Measurement Model
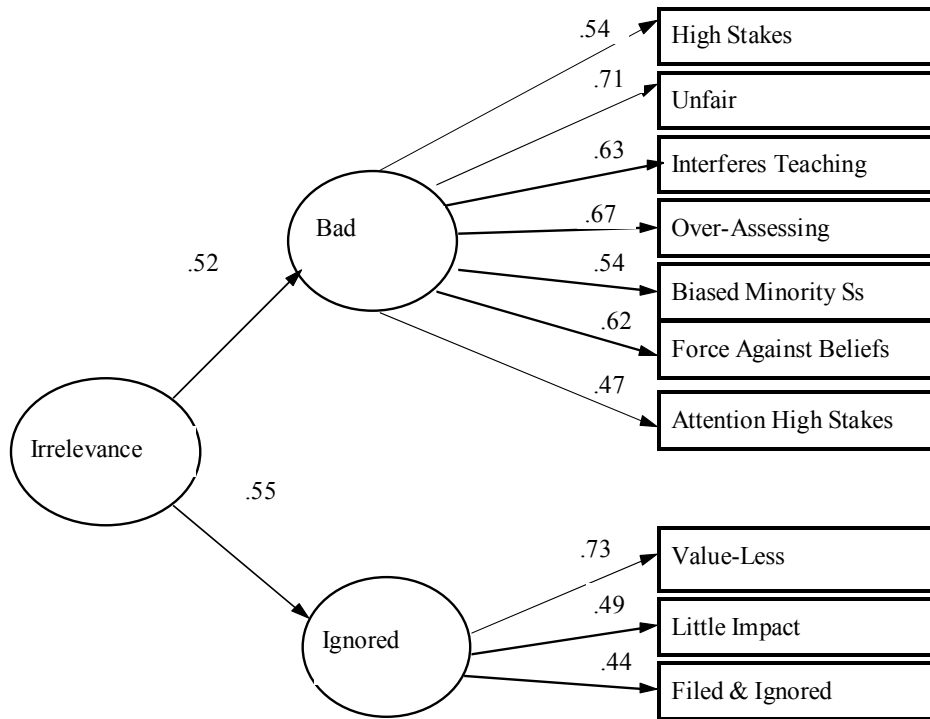
Low sub-scale means indicated that teachers tended to disagree with the conception sub-scale, while high values indicated agreement with a conception. Average item scores for each factor ranged between 2.28 and 4.15. Six of the sub-scales ranged between 3.75 and 4.25 indicating moderate agreement (i.e., all four of the accountability sub-scales and the improvement sub-scales of improve teaching and describe student achievement). The improve student learning sub-scale fell between slightly and moderately agree (3.54), along with the irrelevance sub-scale of bad for teaching (3.40). The irrelevance sub-scale used but ignored was close to disagree (2.28), indicating a reasonably high rejection of the concept that teachers use but ignore assessment.

Because of the small sample size it was not possible to analyse all the data in one all encompassing measurement model. However, it was feasible to examine the factor structure of the sub-scale factor scores to begin to understand how the various

conceptions relate to each other. Thus, a two factor maximum likelihood with oblimin rotation factor analysis was obtained using the ten sub-scale scores (Table 20). Inter-factor correlation was .13, indicating a very small common element between factors. Although, only two factors were identified, it was clear that the *Used But Ignored* factor loaded equally poorly on both factors suggesting a possible third factor that may have been detected had there been more related items.

A possible interpretation of the first factor is largely an improvement-focused view of assessment combined with teacher evaluation. Possibly this is the kind of assessment that participants believed should be used to evaluate teachers for the purpose of improving the quality of diagnosis and students' learning. The second factor loaded largely on the accountability and irrelevance conceptions of assessment suggesting that accountability assessment was irrelevant. However, the high loading of *Evaluate Teachers* with improvement suggested that irrelevance is not equated with accountability conceptions per se. In addition, the high loading of *Improve Teaching* on the same factor as accountability and irrelevance is unexpected, though it suggests that the use of assessment to improve the quality of teaching is irrelevant.

Table 20. CoA-II First Order Factor Loadings

First Order Factors	Factor	
	1	2
Diagnose Ability	**.91**	-.04
Improves Learning	**.89**	-.15
Quality Validity	**.75**	-.01
Evaluate Teachers	**.71**	.31
Used but Ignored	-.14	-.11
Evaluate Learning Objectives	-.04	**.77**
Evaluate Schools	.21	**.69**
Evaluate Students	-.08	**.66**
Bad for Teaching	-.50	**.63**
Improve Teaching	.26	**.56**

Note: Inter-factor correlation $r = .128$

Concluding Comment

This study showed that three measurement models called accountability, improvement, and irrelevance could be identified from the responses of education students and practising teachers to the CoA-II inventory and that the conceptions had two levels and were multi-dimensional. However, data from a larger sample of participants is necessary to effectively map all the measurement models into one integrated model rather than rely simply on a factor analysis of sub-scale scores. Furthermore, more items were required to flesh out four factors, such as *Evaluate Schools* and *Used but Ignored*, which had three or less items. In addition, a large sample of practising teachers was required in order to establish the conceptions of teachers, rather than those of students or trainees.

Study 5: Teachers' Conceptions of Assessment (CoA-III)

Study 4 established three measurement models (CoA-II) (i.e., one each for improvement, irrelevance, and accountability) of teachers' and education students' conceptions of assessment. However, several issues were left unresolved. First, data from a large sample of practising primary school teachers was necessary to establish one integrated measurement model of teachers' conceptions of assessment. Second, the structure of the accountability conception was not well specified through low number of items and the possibility that the student and school dimensions of accountability may be separate conceptions.

Study 5 reports the results from a large-scale survey collection of CoA-III inventory data and establishes a four-factor model of teachers' conceptions of assessment. This study answers some of the questions posed at the end of the literature review. Specifically, a self-report attitude inventory about teachers' conceptions of assessment is developed (CoA-III) and the adequacy of the four-factor model is evaluated. Further, the questionnaire provided information about teacher demographic characteristics (i.e., type and length of pre-service training, length of service, gender, ethnicity, type of school worked in, role in school, and types of assessment training undergone). From each participant's self-identified school, key school-level demographic information was derived (i.e., size, type, and socio-economic status of school, community population size, and ethnic mixture of student population) from the Ministry of Education schools database. How teachers' conceptions of assessment relate to teachers' conceptions of curriculum, learning, teaching, and teacher efficacy are reported in the next chapter as part of Study 6.

Instruments

Several instruments were used in this study. A third version of the

Conceptions of Assessment inventory (CoA-III) was developed. It has been argued

that teachers' practices of assessment, the types of assessment instruments they use,

and the quality of assessment training they have would impact on their conceptions of

assessment (McMillan, 2001b; Quilter, 1998). Therefore, a questionnaire was used to

elicit demographic information, including the amount of training they have had in

assessment, about teachers. A checklist of assessment formats or types was

developed to test whether style of assessment influenced CoA. In addition, a self-

reported inventory of frequency of using different kinds of assessment (i.e.,

assessment practices) was developed.

CoA-III

The 46 items that formed the CoA-II measurement models of Study 4 were

included in the CoA-III Study 5 inventory. The accountability measurement model in

Study 4 had been found not to have good fit to the data, partly because three of the

factors had only three or fewer items. Likewise, the *Used but Ignored* factor in the

irrelevance measurement model only had three items. Thus, a total of 19 additional

items were written to provide more items for factors identified in trial 2 and to

address further conceptions. Specifically, items were written for assessment as a

means of school and teacher accountability (seven items), student accountability (two

items), that assessment is inaccurate (four items), and that assessment is used but

ignored (two items). Further reading of the literature identified, within the context of

the irrelevance conception, the use of assessment as a way to control students or

classes (Torrance & Pryor, 1998) and so four items were written to measure that

115

construct. The resulting 65 items were arranged in the CoA-III so that items relating

to each other were not encountered in groups. The focus of this study is on the CoA-

III, and so the other instruments used in the questionnaire are described in Study 6.

In order to reduce participant workload, a planned missing data design was

used (Graham, Taylor, & Cumsille, 2001). This meant that each participant

completed one of two questionnaire forms. Each questionnaire contained the 65

statement Conceptions of Assessment (CoA-III) Inventory. One questionnaire asked

teachers the frequency of their use of certain assessment types, while the second

questionnaire contained inventories on teacher conceptions curriculum, learning,

teacher efficacy, and of teaching. Data missing at random, ranging from 1.5-8.5%

($M = 2.93\%$, $SD = 1.20\%$) of each variable, were imputed using SPSS EM procedure.

Demographic Questionnaire

In order to assess whether a specific set of assessment practices influenced the

manner that teachers' conceived assessment, teachers were asked to identify which of

up to eleven kinds of assessment they had in mind as they completed the CoA-III. In

both questionnaires, professionally relevant demographic information about the

teacher (i.e., ethnicity, gender, role and length of teaching experience, type and length

of training, and name of school) was collected. From the school name it was possible

to derive various key demographics (e.g., school SES, school size, school ethnic mix,

community size in which the school was located) from the Ministry of Education

Schools database.

Assessment Format or Types

Assessment is associated with a wide variety of information gathering types or

practices (such as those used in Warren and Nisbet's, 1999, survey of Australian

116

mathematics teachers, i.e., observations, oral tests, practical work, interviews, timed tests, projects, homework, assignments), though most commonly it is associated with testing. In the context of mathematics teaching, it has been shown that teachers use a wide variety of techniques but that teachers tended to give greatest importance to teacher observation and student performances and, not surprisingly, least importance to the use of essays (Adams & Hsu, 1998). Similarly, Australian mathematics teachers reported using observation most, while using assignments and journals least often (Warren & Nisbet, 1999). McMillan, Myran, and Workman (2002) found that although teachers used a wide variety of assessment types, they used objective items most often, but that differences were found between language arts and mathematics teachers. The former used performance assessment and projects as much as the objective items, while the latter used teacher-made and publisher-supplied tests as much as objective items.

In contrast, Senk, Beckmann, and Thompson (1997) found, in their study of assessment practices in American high school mathematics classes, that teachers' assessment practices were dominated by reliance on tests and quizzes, a pattern probably consistent with the impact of qualifications frameworks. Kahn (2000) reported that among the high school English teachers in the United States in her study twice as many points were awarded through multiple-choice assessments than all other written and oral assessment types across the full range of English functions (i.e., literature, writing, grammar, listening, speaking, and vocabulary). This may be related to the relatively narrow and formulaic approach to learning the teachers took in their specification of assessment criteria (i.e., the extended writing task rewarded conformity to a formula rather than degree of persuasiveness).

Thus, although teachers use a wide variety of assessment formats or types, what is unknown is whether thinking of certain assessment types is associated with different conceptions of assessment. It might be expected that teachers who prioritise the use of objective items, publisher-supplied tests, and teacher made tests would give greater prominence to accountability conceptions of assessment over improvement conceptions. In contrast, it may be that teachers who associate assessment with informal observation and oral questioning would have a stronger emphasis on the irrelevance of assessment conceptions.

One way to understand how teachers conceive of assessment is to identify the types of assessment they have in mind while completing the questionnaire. In order to address the issue of whether association assessment with certain assessment types relates to differing assessment conceptions, a section just prior to the CoA-III questionnaire was included that allowed teachers to identify which of up to 11 different assessment practices (i.e., unplanned observation, oral question and answer, planned observation, student written work, student self or peer assessment, conferencing, portfolio/scrapbook, teacher made written test, standardised test, essay test, and 1-3 hour examination) they had in mind as they were completing the CoA-III inventory. Multi-dimensional scaling was used to investigate whether it was meaningful to group assessment types by the pattern of responses.

Assessment Practices

Teachers' assessment practices have been researched in a number of studies. For example, Stamp (1987) found in her study of Australian teacher trainees' opinions of assessment practices that there were three distinct factors. These were (a) discovery of individual student's achievements and progress, (b) use of external

examinations to record academic achievement, and (c) student control and selection

of assessment methods and criteria. The same teacher trainees reported believing that

two major types of assessment practices should be used when measuring student

achievement, that is (a) performance on projects, reports, or practical applications and

(b) tests or examinations (Stamp, 1987). Thus, she found that:

> primary education students [teacher trainees] preferred practices which i)
>
> provided for individualized learning and progress, yet ii) maintained a
>
> traditional-academic approach. They did not favour the use of formative
>
> testing measures in the classroom, and they tended to favour the gathering of
>
> subjective, affective types of information when measuring pupil achievement.
>
> (Stamp, 1987, p. 97)

It appears that, at least in a United States survey of primary and secondary

teachers, that teachers assess students regularly for grading purposes; three fourths of

teachers gave a minor test that counted for grades at least once per week (Cizek,

Fitzgerald, Shawn & Rachor, 1995). Approximately four out of five of those minor

tests were ones teachers created themselves, whereas teachers created major tests only

three out of five times. Publishers supplied the balance of tests used for grading

purposes. Nevertheless, the survey found that, although nearly 90% of teachers used

tests to get grading information, a wide variety of assessment types were used

including informal measures of effort, behaviour, teamwork, answers to in-class

questions, participation in class, homework completion, etc.

McMillan (2001b) surveyed middle and high school teachers about the types

of assessment used and the levels of cognitive processing required by their

assessments. Their varimax principal component analysis of 11 assessment type items

and four levels of cognitive assessment found four type factors (i.e., constructed

response included essays, performance-based, and projects; developer-made assessment; objective assessments; and major exams) and two cognitive factors (i.e., deep understanding, reasoning, and higher order thinking; and recall) (Table 21). Using a 6-point frequency scale (i.e., not at all, very little, some, quite a bit, extensively, and completely) McMillan found that teachers relied on self-designed assessment that required demonstration of understanding, reasoning, and application.

These items were incorporated into a measure of teacher assessment practices along with more items adapted from the Entwistle (n.d.) learning beliefs questionnaire to provide more items on cognitive processing of assessments (Table 21). The same frequency of use response scale was used.

Table 21. Assessment Practices and Processes Statements, Factors and Loadings

Factors and Statements	Loading
Constructed Response Assessments	
Performance-based assessments (e.g., structured observations or ratings of performance such as a speech or paper)	.77
Oral presentations	.74
Projects completed by teams of students	.73
Projects completed by individual students	.70
Essay-type questions	.68
Authentic assessments (e.g., 'real world' tasks)	.52
Assessment Developer Made Assessments	
Assessments designed primarily by yourself	.87
Assessments provided by publishers or supplied to the teacher	-.84
Objective Assessments	
Performance on in-class quizzes	.84
Objective assessments (e.g., multiple-choice, matching, short answer)	.68
Major exams or tests	.90
Deep Cognitive Level	
Assessments that measure student understanding	.91
Assessments that measure student reasoning	.88
Assessments that measure how well students apply what they learn	.81
Surface Cognitive Level	
Assessments that measure student recall	.98
Supplementary Items Based On Entwistle (n.d.)	
Surface Cognitive Level	
Assessments that measure ability to build up knowledge by getting facts and information	—
Deep Cognitive Level	
Assessments that measure whether students see things in a different and more meaningful ways	—
Assessments that measure how students are developing as individuals	—
Assessments that measure whether students can derive abstract principles from ideas or information	—
Assessments that measure student ability to understand relationships between ideas or information	—

Assessment Literacy Training

Assessment literacy, based on the American standards (AFT, NCME, & NEA, 1990), is understood as "the ability to design, select, interpret, and use assessment results appropriately for educational decisions" (Quilter, 1998, p. 4). It is anticipated that teachers with more training in the use of assessment in education would have different conceptions of assessment than those with less training. It seemed possible that teachers with less assessment literacy training would be more likely to conceive

of assessment as irrelevant or disagree with the assessment as improvement

conceptions. Thus, it was decided to investigate the amount of training in assessment

that teachers had participated in.

Research has indicated that the vast majority of teachers have limited

understanding of the qualities of assessment information (e.g., reliability, validity of

inferences, and statistical terminology), whether it be derived from their own

observation evaluation of a student, from a student's external standardised test mark,

or from their own in-class performance assessments (Hambleton & Slater, 1997;

Impara, Divine, Bruce, Liverman, & Gay, 1991; McMillan, Myran, & Workman,

2002; Mertler, 1999; Plake, Impara, & Fager, 1993; Quilter, 1998; Stiggins, 2001).

Gipps, et al. (1995, p. 2) argued that the research on assessment they conducted in

primary schools in the early 1980s "had shown that teachers' understandings of issues

in assessment was very limited; while there was widespread use of standardized tests

of reading and maths, there was little understanding of how the scores were derived,

or what they meant, and no understanding of issues such as reliability and validity".

Furthermore, teachers' personal experience of assessment as positive or

negative while they themselves were students is positively correlated with their

current attitude towards assessment (Green & Stager, 1986; Quilter, 1998). Stamp

(1987) found that teacher trainees who had experienced individual inquiry methods of

assessment in primary and secondary school tended to agree with the use of the same

methods as a classroom assessment practice, while those who had experienced

teacher-centred testing methods in their own schooling tended to agree with the use of

teacher controlled testing. It has also been demonstrated that greater assessment

literacy correlated with more positive attitudes toward classroom assessment (Quilter,

1998; Quilter & Chester, 1998). In addition, it has been found that training in

122

assessment may have very little impact on changing teacher trainees' assessment frameworks (Stamp, 1987).

Thus it was decided to collect information on the amount of assessment literacy training teachers had received. It was not possible to probe more deeply because the research's focus was on conceptions of assessment, not the ability to interpret accurately assessment information. Teachers were asked to indicate, on a non-exclusive basis, what kind of assessment training they had participated in. The options were arranged hierarchically from none, through some hours as part of pre-service training, and 1/2 to 1-day workshop or seminar, to completion of formal courses in assessment at the undergraduate or postgraduate levels.

Participants

CoA-III questionnaires were sent to a national stratified sample of 800 schools, randomly selected by school SES, size, location, and type in proportion to the percentage of student population in each category of school. Questionnaires were addressed to the principal who was asked to call for two volunteer participants from among the teachers of Year 5 to 7 students. Questionnaires were sent out during the last term of the school year. Despite not sending out any follow-up reminders, a total of 525 CoA-III inventories, return rate of 33%, were returned in time for analysis.

Characteristics of individual teachers participating are identified in Table 23. The teachers in the study were for the most part (a) New Zealand European (83%), (b) female (76%), (c) highly experienced with 10 or more years teaching (63%), (d) employed as teachers rather than managers or senior teachers (54%), (e) employed in contributing or full primary schools (89%), and (f) well trained with three or more years training (55%). The demographic characteristics of the individual teachers in this sample reasonably reflect those of the New Zealand teaching population (Table

22) as determined in the 1998 teacher census conducted for the Ministry of Education

(Sturrock, 1999).

Table 22. CoA-III Key Demographic Characteristics Comparison

Characteristic	2001 CoA-III Study	1998 Teacher Census
Sample Size	525	23,694
NZ European	83%	87%
Female	76%	71%
Long Service	63%	49%[a]

Note. [a]This figure averaged for both primary and secondary sectors as separate sector information was not available.

Teachers from 290 schools provided 491 CoA-III questionnaires, while a

further 36 were returned without school identification (Table 24). About one-third of

teachers were employed in low socio-economic status (SES) schools, while just fewer

than 30% of teachers worked in high SES schools. This distribution represented a

very acceptable sampling of the distribution of teachers by school SES. Just over half

of the teachers worked in large urban area schools, representing a proportional return

by school type (i.e., approximately 40% of primary schools are smaller than 120

students and are largely rural). However, this represents a significant over-sampling

of small schools based on student population proportions, since only about 10% of

students are in such small schools.

The proportion of students who are reported to be of New Zealand European

or Pakeha ethnicity was used to group schools into majority or minority ethnic school

(Hattie, 2002). Those schools that had more than 25% of students with New Zealand

European ethnicity were classified as majority, while schools that had up to a

maximum of 25% New Zealand European students on their roll were classified as

minority. Just over three-quarters of participants came from majority ethnicity

schools.

Table 23. CoA-III Participant Characteristics

Characteristics	Gender			Total
	Male	Female	Missing	
Ethnicity				
NZ European/Pakeha	99	335	1	435
NZ Maori	7	28		35
Other	10	21		31
Asian	1	6		7
Pacific Nation	1	5		6
Missing	2	4	5	11
Years Teaching				
More than 10	85	248		399
Between 2 and 5	14	60		74
Between 6 and 10	13	56	1	70
Less than 2	8	33		41
Missing		2	5	7
Role				
Teacher	47	234	1	282
AP or DP	20	65		85
Principal	35	35		70
Senior Teacher	10	54		64
Other	6	5		11
Trainee Teacher		2		2
Missing	2	4	5	11
Years Training				
Less than 1 Year	1	4		5
1 Year	2	13		15
1-2 Years	11	24		35
2 Years	12	36		48
2-3 Years	25	91	1	117
3 Years	34	308		142
More than 3 Years	34	113		147
Missing	1	10	5	16
School Type				
Contributing Primary	47	182	1	230
Full Primary	61	175		236
Intermediate	9	39		48
Secondary	1			1
Missing	2	3	5	10
Type of Training				
Early Childhood	1	2		3
Primary	110	375	1	486
Secondary	3	6		9
Both Primary and Secondary	6	12		18
Missing		4	5	9
Total	120	399	6	525

Table 24. CoA-III Participants by School Characteristics

Characteristic	Frequency	Percent
Socio-Economic Status (Decile)		
Low		
1	66	12.6
2	54	10.3
3	53	10.1
Middle		
4	48	9.1
5	26	5.0
6	47	9.0
7	42	8.0
High		
8	58	11.0
9	48	9.1
10	48	9.1
Missing	35	6.7
Community Population Type		
Urban		
Main Urban	275	52.4
Secondary Urban	26	5.0
Rural		
Minor Urban	53	10.1
Rural	133	25.3
Missing	38	7.2
School Size		
Large (>350)	145	27.6
Medium (121-350)	195	37.1
Small (<=120)	150	28.6
Missing	35	6.7
School Ethnic Mix		
Majority (>26% European)	403	76.8
Minority (<=25%)	87	16.6
Missing	35	6.7
Total	525	100.0

Thus, data in this study were from a relatively homogenous population of full and contributing primary school teachers, largely representative of the New Zealand population, except for an over-representation of teachers in small schools. It is important to test if participants would have different CoA scale scores according to these school or personal demographic factors.

Results

In contrast to previous trials where insufficient participants precluded development of an overarching measurement model, the larger sample size available in this study permitted determination of the relationship between the various conceptions. SEM of the CoA-III data allowed integration of the various CoA-II measurement models developed in Study 4 into one comprehensive model of teachers' conceptions of assessment.

CoA-III Measurement Model

The process of creating a meaningful, well-fitting model of teachers' conceptions of assessment with the CoA-III data removed 15 items. Items were removed that caused negative error variance by being overly correlated with each other or which had low loadings on their intended factors. The remaining 50 items in the CoA-III resulted in a well fitting measurement model (χ^2 = 3217.68; *df* = 1162; *RMSEA* = .058; *TLI* = .967) containing four correlated major factors, which constitute the conceptions of irrelevance, improvement, school accountability, and student accountability (Figure 6). The first two factors are second-order purposes that have three or four first-order factors, while the latter two are stand-alone first-order factors.

The first-order factor loadings for irrelevance clearly indicate that the irrelevance conception consists of three conceptions; specifically, assessment is bad for teaching (λ = .91), teachers may use assessment but they ignore it (λ = .74), and that assessment is inaccurate (λ = .41). The first-order factor loadings for improvement consist of four conceptions. These are that assessment improves student learning (λ = .92) and teaching (λ = .86), that assessment describes student abilities, knowledge, and thinking (λ = .93), and that assessment information is valid because

of its dependability (λ = .80). The school accountability factor is made up of six

statements that focus on using assessment to evaluate the worth of schools; while the

student accountability is formed by 7 statements that focus on using assessment to

evaluate student progress against achievement objectives or to make placement or

selection decisions about students.

Correlations between the four factors help to clarify the nature of teachers'

conception of assessment (Table 25). The irrelevance factor was quite highly but

inversely correlated with the Improvement conception, suggesting there may be a

bipolar relationship, but there was sufficient variance not accounted for to leave these

as two independent factors. The irrelevance factor has a zero correlation with school

accountability conception (r = -.12; p = .1296) with α = .01, and is moderately

correlated with student accountability. The improvement factor is quite highly

correlated with school accountability and moderately correlated with student

accountability. The two accountability factors are moderately correlated with each

other. It would appear that Crooks' (1990) call for school-based self-evaluation as a

means of determining school effectiveness has been heeded and implemented in the

subsequent decade.

Table 25. Intercorrelations CoA-III Model of Conceptions of Assessment

Factor	1	2	3	4
1. Improvement	—	-.69	.58	.32
2. Irrelevance		—	-.12	.29
3. School Accountability			—	.58
4. Student Accountability				—

Figure 6. CoA-III Measurement Model of Conceptions of Assessment
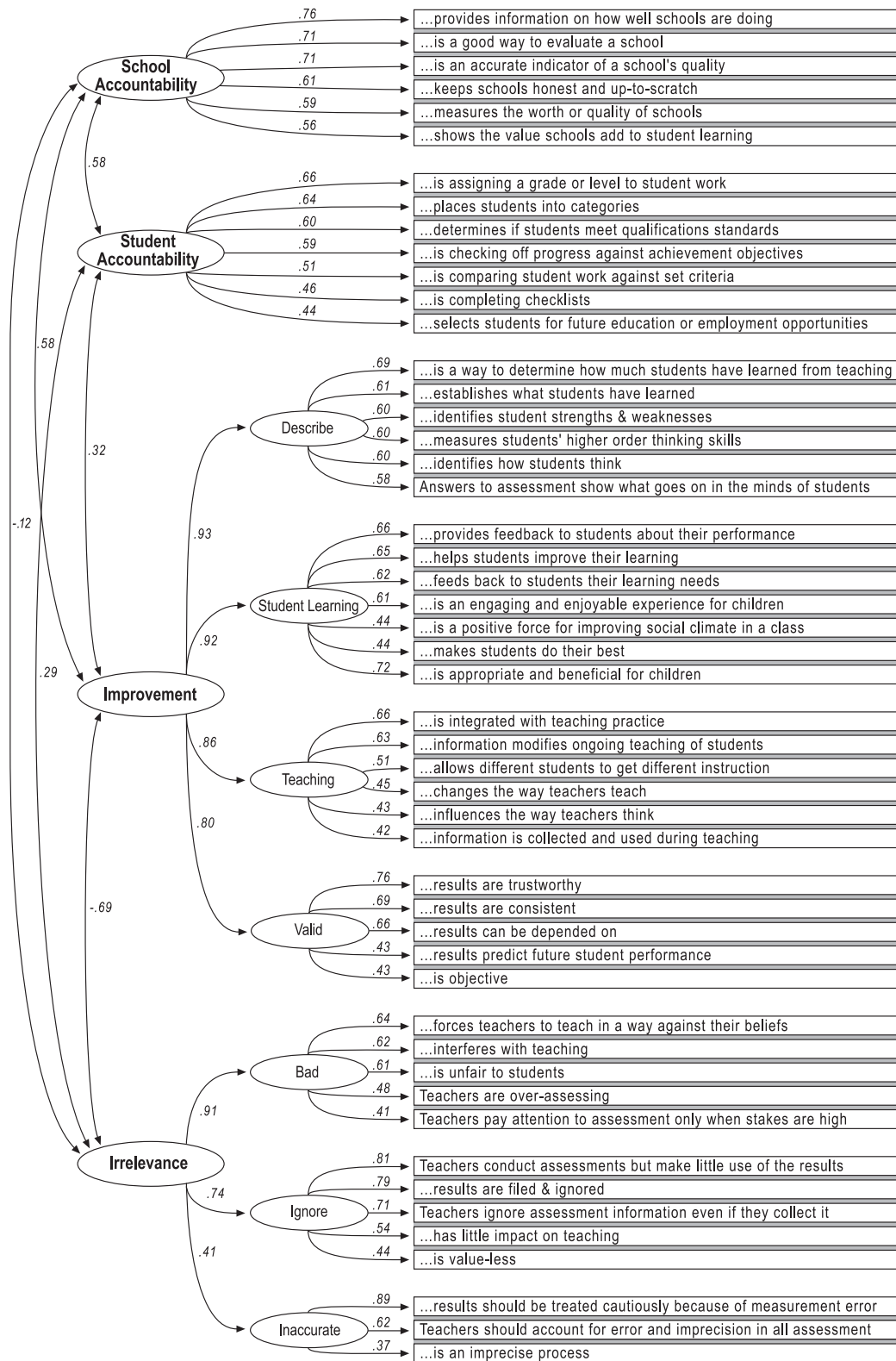
Differences in CoA-III Mean Scale Scores

Mean scale scores were calculated by dividing total scale score by the number of items in each scale (Table 26). This was done so that scale scores could be interpreted by the score anchor terms and compared to each other. There are a total of nine sub-scales based on 50 statements. Table 22 shows the number of items, the average scale score and standard deviation, the loading of each scale on applicable second-order factor, the scale internal reliability, and the average loading of statements for each scale. Based on the average item loadings onto first-order factors (range .52-.66), there is a strong indication that the items are related to each other as a factor. In addition, the scale factors exhibit acceptable to good internal reliabilities (range .63-.81). The two second-order factors are well explained by the first-order factor with loadings ranging from .41 to .93. Variance in mean scale scores ranged between 12-15% of the maximum scale score of 6. These loadings, reliabilities, and variances contribute to the overall quality of the measurement model, as indicated by the fit indices values reported above.

Table 26. CoA-III Scale Characteristics

Second-Order Factors First-Order Factors	# of Items	Mean Scale Score (*SD*)	Loading on Second-Order Factor	Scale Reliability (α)	Mean Item Loading on First-Order Factor
Improvement- Information					
Improve Teaching	6	4.32 (.71)	.86	.68	.52
Improve Learning	7	3.85 (.76)	.92	.79	.59
Quality Validity	5	3.32 (.81)	.80	.73	.59
Describe Ability	6	3.87 (.76)	.93	.78	.61
Irrelevance					
Bad for Teaching	5	2.73 (.77)	.91	.68	.55
Used but Ignored	5	2.27 (.81)	.74	.78	.66
Inaccurate	3	3.86 (.93)	.41	.63	.63
School Accountability	6	3.53 (.81)	—	.81	.66
Student Accountability	7	2.85 (.78)	—	.75	.56
Total CoA	50	3.42 (.74)	—	.85	—

The highest mean scale score (i.e., just over moderately agree) is on Improve

Teaching, followed by three scales just under moderately agree, (i.e., Improve Student

Learning, Describe Student Abilities, and Inaccurate Assessment is Irrelevant). The

lowest scale score (i.e., just above mostly disagree) was for Used but Ignored, followed

by two scales just under slightly agree (i.e., Bad for Teaching and Student

Accountability). For each scale, it is interesting to investigate the number of teachers

who agree strongly or weakly and who disagree with each concept (Table 27 & Figure

7). Scale scores in the range 1.00 to 2.99 were considered disagreement, while those

from 3.00 to 4.50 were considered lower agreement and those between 4.51 and 6.00

were treated as strong agreement.

Table 27. Percentage of Teachers by Agreement Level for CoA-III Scales

Conception of Assessment	Disagree (%)	Low Agree (%)	Strongly Agree (%)
Accountability Students	24.4	65.1	10.5
Accountability Schools	53.3	43.8	2.9
Improve Describe	10.7	72.0	17.3
Improve Student Learning	11.8	65.5	22.7
Improve Teaching	2.9	59.6	37.5
Improve Valid	30.1	62.9	7.0
Irrelevance Bad	62.5	35.8	1.7
Irrelevance Ignore	79.2	20.0	0.8
Irrelevance Inaccurate	14.3	59.2	26.5

Four conceptions were strongly agreed with by more than a quarter of teachers

(i.e., improve teaching—37.5%; irrelevance inaccurate—26.5%; and the inverse of

irrelevance bad—62.5% and irrelevance ignore—79.2%). Three assessment

conceptions were disagreed with by more than a quarter of teachers (i.e., accountability

schools—53.3%; improve student learning—44.8%; and improve valid—30.1%).

Clearly, most teachers agree strongly that assessment is not irrelevant, about half of

teachers disagreed that assessment improves student learning and makes schools

accountable, while a third of teachers disagreed with the validity of assessment and

another third agreed strongly that assessment would improve teaching, and a quarter of

teachers agreed strongly that assessment was inaccurate. Although a good shorthand

summary of New Zealand teachers' conceptions of assessment this analysis fails to

identify how the conceptions relate to each other. Nevertheless, it is clear that pockets

of New Zealand teachers have strongly held and conflicting conceptions of assessment.

Figure 7. Percentage of Teachers by Agreement Level for CoA-III Scales
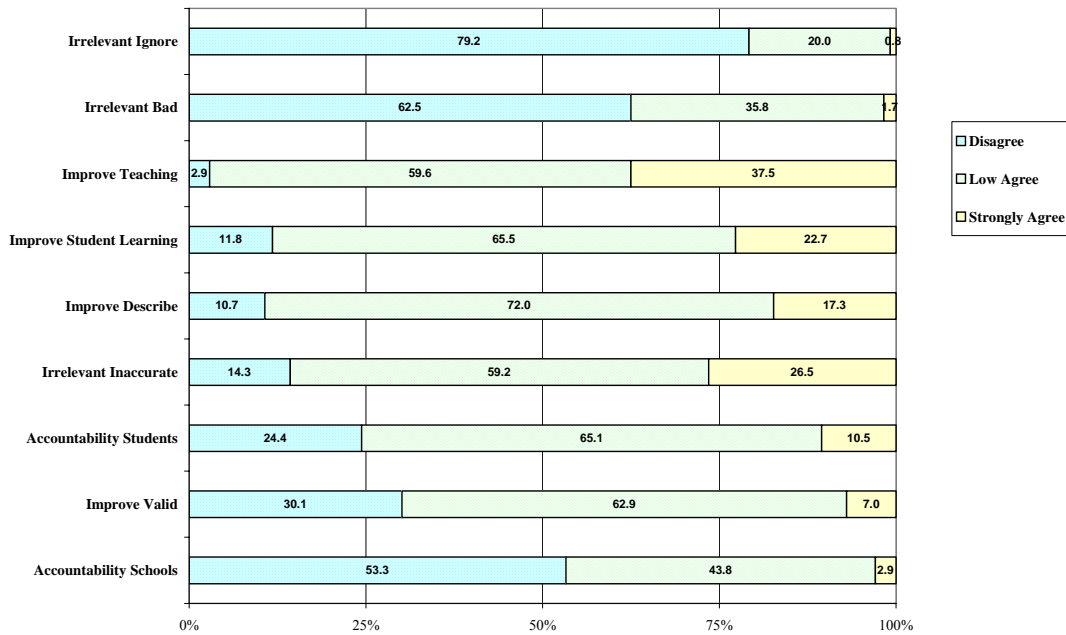
Teacher & School Characteristics

Teacher characteristics (outlined in Table 23) were examined as a possible

source of variance in COA-III subscale values. Because the overwhelming majority of

teachers were NZ Europeans, trained as primary teachers, and worked in composite or

full primary schools it was decided not to analyse these factors. Since there were a

large number of response categories for teacher role, teacher years of experience, and

years of training, those variables were collapsed into fewer approximately equal-sized

categories. Teacher role was collapsed to teacher ($n = 281$) and manager or leader

($n = 218$); years of experience was collapsed to ten years or less ($n = 180$) and more

132

than ten years (*n* = 319), while years of teacher training was collapsed into four

categories; less than two years (*n* = 107), two to three years (*n* = 116), three years

(*n* = 136), and more than three years (*n* = 140).

Multivariate *ANOVA*s found no statistical difference for any of the COA-III

subscales for teacher gender, teacher years of training, or teacher years of experience.

The *F* tests for teacher roles, based on the linearly independent pair-wise comparisons

among the estimated marginal means, found one statistically significant difference,

$F(1,501) = 11.691$, $p = .001$, for only the improve student learning subscale. Managers

and leaders (*M* = 4.01) agreed more strongly that improving student learning defined

assessment than teachers (*M* = 3.74). This particular distinction in attitude between

leaders and teachers, though it may be an artefact of experiment-wise statistical testing,

has been found in the literature on teachers' implicit theories about teaching (Clark &

Peterson, 1986).

That only one of the nine COA-III subscales had statistically significant

difference suggested that differences in role are not powerful in shaping teachers'

conceptions of assessment. Thus, other than for one subscale (i.e., improve student

learning by role in school), teacher gender, years of training, years of experience, and

role in school were irrelevant to mean scale scores for the teachers' conceptions of

assessment inventory. This indicated the relative stability and generalisability of

teachers' COA-III conceptions.

The characteristics of the schools in which the participants worked (Table 24)

were examined to determine whether mean COA-III scale scores would be affected by

school size, school SES, community population size, or ethnic mixture of student

population. To permit analysis of reasonably comparable cell sizes, school SES was

collapsed into three categories (i.e., low, medium, and high) and school community

population size was collapsed into two categories (i.e., urban and rural). Multivariate *ANOVAs* of mean CoA-III scale scores found no statistically significant differences (i.e., School SES, $F(18, 434) = 1.207$, Wilks' $\lambda = .947$, $p = .248$; School Size, $F(18, 434) = 1.047$, Wilks' $\lambda = .954$, $p = .403$; Community Type, $F(18, 434) = 1.064$, Wilks' $\lambda = .976$, $p = .389$) for main or interaction effects. Thus, school characteristics did not differentiate in a statistically significant way the mean scores for the nine teachers' conceptions of assessment factors.

Further evidence for the credibility of this model can be found by inspecting the means for the various conceptions across the three studies. If the model generation process has not generated chance results as a consequence of population characteristics, then the mean scale scores should be stable across populations in the three studies conducted to develop this instrument. Factors were matched across the three studies and where factors were not closely labelled they were assigned to the conception in CoA-III to which they were most similar. When more than one factor in earlier studies matched a factor of CoA-III they were averaged.

Mean scale scores across the three studies were quite similar (Table 28) with low absolute mean difference between study scale score and average for all three studies. However, three of the factors (i.e., *Accountability Students, Irrelevance Bad,* and *Irrelevance Ignore*) showed much larger absolute discrepancies in mean scores across the three studies (between 10 and 17% of mean scale score). Further, it should be noted that the average mean scale score was lowest in the CoA-III. In addition, it should be noted that the average of all scale scores was approximately halfway between slightly agree and moderately agree. Thus, there is evidence that teachers' conceptions of assessment have remained relatively stable despite changes in items and different populations (i.e., trainee teachers and practising teachers).

134

Table 28. CoA Mean Scale Scores Across Studies 3-5

Conception of Assessment	CoA-I	CoA-II	CoA-III	*Average*	Absolute Mean Difference
Accountability Schools	3.81	3.95	3.53	*3.76*	0.16
Accountability Students	—	4.02	2.85	*3.43*	0.59
Improve Teaching	3.95	4.10	4.32	*4.12*	0.13
Improve Learning	3.56	3.54	3.85	*3.65*	0.13
Improve Valid	3.72	3.20	3.32	*3.41*	0.20
Improve Describe	3.87	3.72	3.87	*3.82*	0.07
Irrelevance Bad	3.60	3.40	2.73	*3.24*	0.34
Irrelevance Ignore	3.01	2.28	2.27	*2.52*	0.33
Irrelevance Inaccurate	3.45	—	3.86	*3.66*	0.21
Average	*3.62*	*3.53*	*3.40*	*3.51*	0.08

Intercorrelations between CoA scale scores across the three studies were also reasonably high, with greatest correlation between CoA-I and CoA-II (Table 29). Those two studies were predominated by teacher trainees and undergraduate education students, while CoA-III consisted of practising teachers. The evidence from the pattern of means and intercorrelations supports the assertion that the four meta-factor model of teachers' conceptions of assessment is stable across populations.

Table 29. Intercorrelations CoA-III Mean Scale Scores Across Studies

Study	1	2	3	4
1. CoA-I	—	.93	.73	.89
2. CoA-II		—	.68	.89
3. CoA-III			—	.92
4. Average				—

Assessment Format or Types

In order to simplify analysis, multidimensional scaling (MDS) was used to group assessment types. MDS or perceptual mapping is a series of techniques that identify key dimensions underlying respondent reactions to objects (Hair, Anderson, Tatham, & Black, 1998). Objects are mapped into multidimensional space based on their distance from each other. The resulting spatial map shows the relative position of

objects from each other. To establish the quality of an MDS solution, it is

recommended that there be $d*4$ objects for d dimensions, the proportion of the variance

of the disparities not accounted for the MDS model (i.e., Kruskal's stress) be less than

.05, and that the proportion of variance (R^2) explained be greater than .60 (Hair et al.,

1998).

MDS, using Alscal Euclidean distances procedure, reduced the 11 types of

assessment to four meaningful dimensions, with good fit characteristics (Kruskal's

stress = .026; R^2 = .997) (Table 30). Although this was more dimensions than is

normally recommended for the number of objects, the two dimension solution did not

reach the .05 threshold for Kruskal's stress and the three dimension solution did not

generate as meaningful a pattern of results as the four dimension solution. Five types of

assessment grouped together and were labelled teacher controlled classroom

assessments (i.e., teacher made written tests, standardised tests, student written work,

planned observations, and student self or peer assessments). Two formal examination

types grouped together (i.e., 1-3 hour exams and essay tests), while three oral

assessments grouped together (i.e., conferencing, oral question and answer, and

unplanned observations). The portfolio method stood by itself.

Table 30. Multidimensional Scaling of Types of Assessment

Assessment Types	Dimension	Dimension Coordinates			
		1	2	3	4
1-3 Hour Exams	Exam	-3.70	.13	-.28	.09
Essay Test	Exam	-3.43	.12	-.24	.05
Conference	Oral	.44	.09	1.05	.73
Oral Question and Answer	Oral	.57	1.01	.52	.22
Unplanned Observation	Oral	.81	.93	.21	-.87
Portfolio/Scrapbook	Portfolio	-.15	-1.62	.61	-.52
Planned Observation	Teacher	1.35	-.04	-.06	-.01
Self or Peer Assessment	Teacher	1.05	-.36	.27	.40
Standardised Tests	Teacher	.58	-.21	-1.35	.27
Student Written Work	Teacher	1.41	.11	-.13	.08
Teacher Made Written Tests	Teacher	1.06	-.16	-.57	-.45

Figure 8 shows the four types of assessment dimensions mapped in three-dimensional space with the axes rotated in such a way that the clustering of types was more visible. The five teacher controlled types are marked T; the three oral types are marked O; the two exam types are marked E, and the portfolio type is marked P.

Figure 8. Three Dimensional Mapping of MDS Assessment Type Dimensions
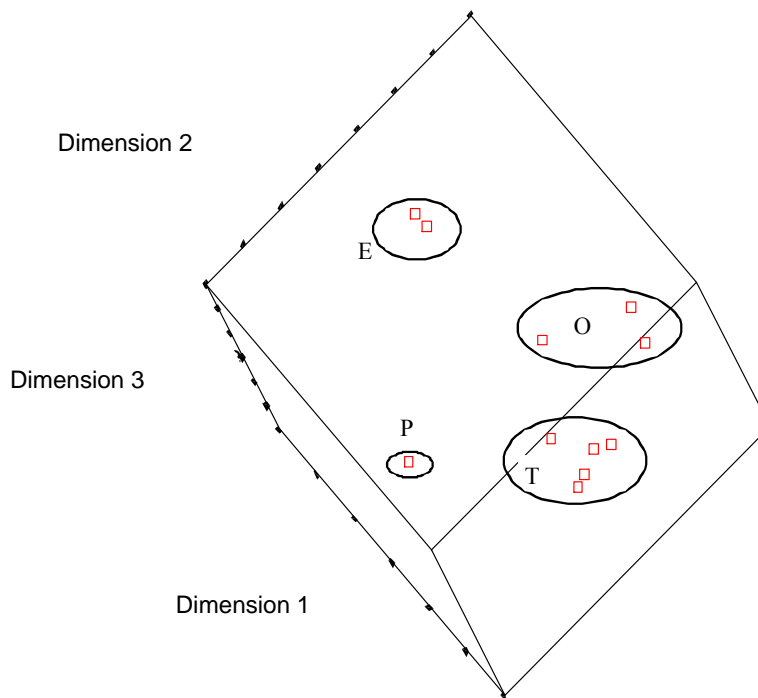
A frequency score for each assessment dimension was calculated tallying the number of times each type had been selected by each teacher (Table 31). Very few teachers selected no oral or teacher controlled types, or any of the exam types, while two-fifths did not select portfolio type. On average teachers selected four of the five teacher controlled assessment types, one and a half of the three oral types, and just six in ten selected the portfolio type. This is in stark contrast to the 61 teachers that thought of either one or both of the exam type assessments.

Table 31. Frequency of MDS Assessment Types Scores

	Score						
Assessment Dimension	0	1	2	3	4	5	Average
Oral	65	72	112	266	—	—	1.55
Exam	454	51	10	—	—	—	.14
Teacher	46	6	16	35	128	284	4.03
Portfolio	204	311	—	—	—	—	.60

Multivariate analysis of variance tests revealed no statistically significant differences in mean scores for the nine Conceptions of Assessment scales for main or two-way interaction effects of the four assessment type dimensions (Table 32).

Table 32. Multivariate Results COA-III by Assessment Type

Effect	Wilks' λ	*F*	*df*	*P*
Oral	.95	.83	27	.72
Exam	.94	1.46	18	.10
Teacher	.90	1.07	45	.34
Portfolio	.98	.89	9	.54
Oral * Exam	.92	1.13	36	.27
Oral * Teacher	.81	.91	108	.73
Oral * Portfolio	.93	1.20	27	.22
Exam * Teacher	.91	1.25	36	.14
Exam * Portfolio	.98	.58	18	.91
Teacher * Portfolio	.92	1.12	36	.29

Thus, it was apparent that, regardless of the type of assessment contextualising teacher responses to the COA-III questionnaire, mean scores for teacher assessment conceptions were consistent across type dimensions. In other words, teachers' conceptions of assessment were general and constant whatever kind or type of

assessment teachers had in mind. Basically, assessment type was irrelevant to teachers' conceptions of assessment.

Assessment Practices

This study makes a distinction between the type or format of assessment associated with the conception of assessment and the frequency with which teachers practice or use differing types of assessment. Just over 230 teachers responded to the frequency of practices of assessment questions. A full exploratory and confirmatory factor analysis was carried out on the assessment practices (i.e., types and cognitive processing) statements simultaneously (unlike the McMillan (2001b) study). A three factor structure was detected on the basis that at least three items had to load $>.30$ each on each factor. Items that loaded poorly or did not load logically on any factor were dropped. The subsequent result was tested with measurement modelling (Figure 9). Overall fit of the model to the data was excellent ($\chi^2 = 216.152$; $df = 51$; $TLI = .986$, $RMSEA = .047$). Intercorrelations between assessment practices scales were weak to moderate (Table 33). The deep and informal practices loaded most strongly on the assessment practices model.

Table 33. Intercorrelations Assessment Practices Subscales

Subscale	1	2	3
1. Deep	—		
2. Informal	.38**	—	
3. Formal	.16*	.24**	—

Note. *p<.05; **p<.01

Figure 9. Assessment Practices Measurement Model
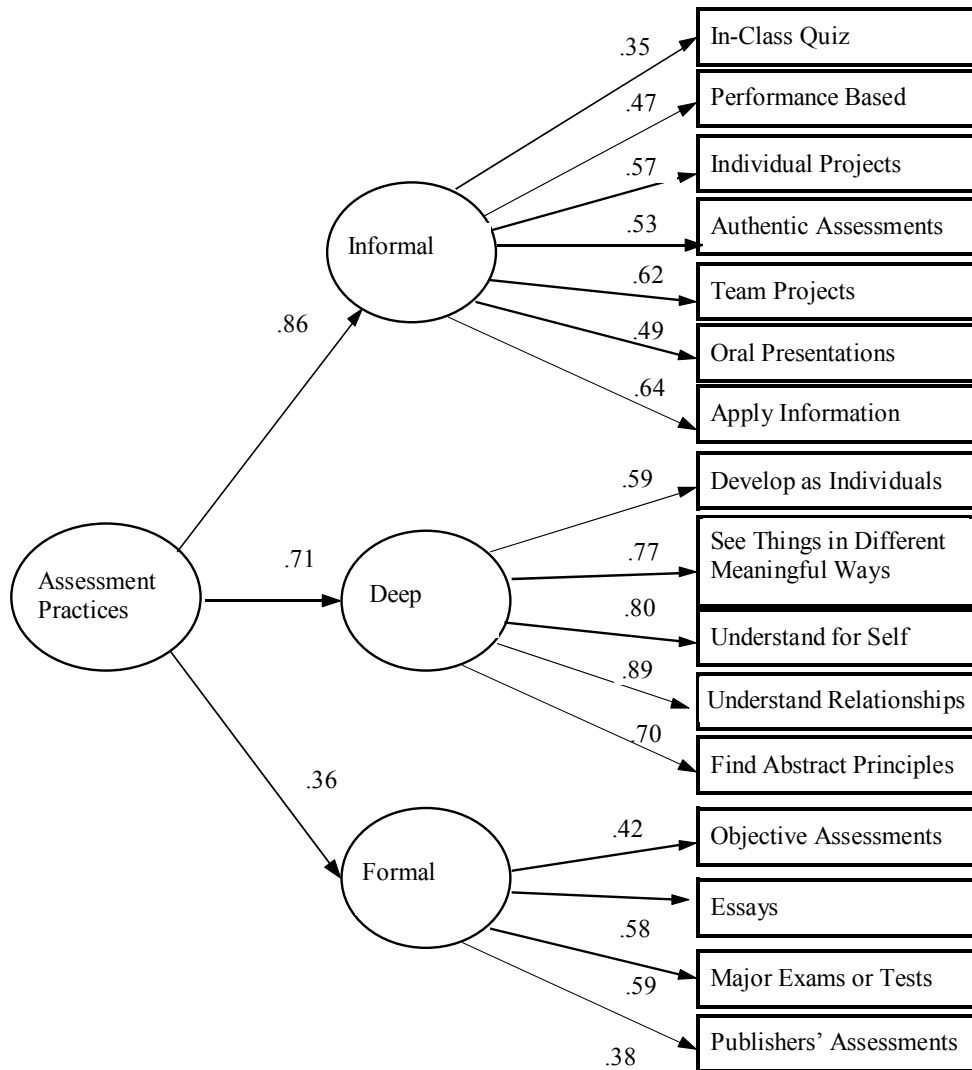
The three factors detected were use of deep cognitive processing ($\lambda = .71$),

informal classroom assessment types ($\lambda = .86$), and formal assessment types ($\lambda = .36$).

Use of deep cognitive processing assessments practices related to requiring relational or

abstract thinking, seeing knowledge in new ways, developing as an individual, or

understanding material for one self. Note that all these items are derived from the deep

processing items in Entwistle's (n.d.) learning conceptions questionnaire. The mean

scale score for the deep factor was 3.34 (*SD* = .80), with acceptable reliability of

$\alpha = 86$. Informal practices related to items that in McMillan's (2001b) study had been

two factors of constructed response and teacher-made assessments. These assessments were those that might be used by a teacher in classroom environments (i.e., projects, quizzes, performance or authentic assessments, and presentations). The mean scale score for the informal factor was 3.68 (*SD* = .56), with acceptable reliability of α = 72.

The formal assessment practices related to those associated with objective, essay, or examination assessments. This factor parallels the McMillan (2001b) objective assessment type factor, although they located essays in the constructed response factor while this study located that type in formal assessment. The mean scale score for the formal factor was 2.57 (*SD* = .60), with low estimate of reliability, α = 54. Unlike McMillan (2001b), despite one more surface type statement being available, no surface assessment practices factor was detected. It may be that the formal assessment type is considered to be surface among this group of teachers.

Thus, teachers reported using deep and informal assessments between some and quite a bit of the time, while their use of formal assessment was between 'very little' and 'some of the time'. This latter mean is lower than the McMillan (2001) mean of secondary teachers for objective assessments of between 'some' and 'quite a bit' and the mean for deep assessments is lower than that study's mean for reasoning assessments of 'quite a bit'. The informal assessment mean is similar to the McMillan (2001) mean for performance assessments of between 'some' and 'quite a bit'.

The intercorrelations between the assessment practices and the conceptions of assessment scales ranged from nil to weak (Table 34) with only two correlations approaching or exceeding .30. The deep practices correlated most strongly with improvement of student learning and teaching scales. On the other hand, the informal classroom related practices loaded on three of the four improvement scales (except for valid) and both the accountability conceptions. The formal practices correlated

141

strongest with the accountability of student learning conception. Except for a very weak negative correlation of deep to inaccurate, none of the practices correlated with any of the irrelevance scales. Only the formal practices scale correlated with the valid scale, suggesting that teachers perceived that only formal assessments meet the validity criteria.

Table 34. Intercorrelations Assessment Practices and CoA-III

	Assessment Practices		
Conception of Assessment	Deep	Informal	Formal
Accountability School System	.16*	.21**	.20**
Accountability Students	.08	.22**	.26**
Improve Describe Students	.19**	.30**	.24**
Improve Student Learning	.22**	.32**	.21**
Improve Teaching	.27**	.23**	.08
Improve Valid	.08	.11	.16*
Irrelevance Bad	-.08	-.02	.03
Irrelevance Ignore	-.12	-.06	-.01
Irrelevance Inaccurate	-.03*	.08	-.03

Note. * p<.05; **p<.01

In order to examine more closely the effect of assessment practices on conception of assessment, scores were classified into low, middle, and high categories. Scores more than one standard deviation below the mean were classified as low, those one standard deviation above the mean were classified as high, and those within one standard deviation of the mean were classified as middle. For deep assessment practices 41 teachers were low, 150 were middle, and 34 were high; for informal assessment practices 34 were low, 158 were middle, and 33 were high; and for formal assessment practices 24 were low, 163 were middle, and 38 were low.

Multivariate F tests of the effect of assessment practices (Deep, $F(18, 396) = 1.272$, Wilks' $\lambda = .894$, $p = .202$; Informal, $F(18, 396) = 1.268$, Wilks' $\lambda = .894$, $p = .205$; Formal, $F(18, 396) = .477$, Wilks' $\lambda = .958$, $p = .967$) based on linearly

independent pair-wise comparisons among the estimated marginal means of those three groups found no statistically significant differences in COA-III subscales.

Assessment Literacy Training

The participants were asked to select all categories of assessment literacy training that applied to them (Table 35). About one in seven teachers noted that they had had no training in assessment, while a third had received some hours on assessment as part of their pre-service teacher training. Nearly half had attended a half to full day workshop or seminar on assessment at some time in their service as teachers. This is somewhat greater than the 34% of teachers who reported participating in assessment professional development in the period March 1997-March 1998 (Sturrock, 1999). Just over one in five had completed an undergraduate paper or course on assessment, while one in twenty had finished a postgraduate paper or course. This value is not significantly different to the 26% of teachers in years 0 to 5 who, in a survey of use of diagnostic assessment tools in literacy and numeracy, reported participating in tertiary courses including some assessment component since pre-service training (Croft, Strafford, & Mapa, 2002).

Table 35. CoA-III Participant Assessment Literacy Training

| Amount and Type of Training | Gender | | | Total |
	Male	Female	Missing	(% of Participants[a])
None	12	62	2	76 (14%)
Some Hours in Pre-Service Training	37	146		183 (35%)
Half to 1 day Workshop or Seminar	51	168	1	220 (42%)
Completed Undergraduate Paper	32	82		114 (22%)
Completed Postgraduate Paper	9	19	1	29 (6%)
Other	27	84	1	111 (21%)
Subtotal[b]	168	560	5	733

Note. [a]Percentage calculated against total of 525 participants.
[b]Total exceeds 525 as participants were instructed to select all that apply.

In addition, teachers supplied alternative other types of assessment training that they had participated in. Of the one in five who had received training in assessment through other means, four categories accounted for just over three quarters of all alternative methods ($n = 86$). These included participation in Ministry of Education funded assessment improvement contracts such as Assessment for Better Learning ($n = 33$), school-based in-service courses ($n = 29$), attendance at short courses ($n = 11$), and components of courses ($n = 13$). Because of the wide diversity of low frequency response categories offered by teachers these data were not utilised in the analysis.

Generally, very few participants had received any extensive formal course work in assessment, while in-service workshops and pre-service lectures accounted for the bulk of assessment literacy training. This overall lack of assessment literacy training is consistent with international trends (Black & Wiliam, 1998; Plake & Impara, 1997; Stiggins, 2001).

Multivariate *ANOVA*s revealed no statistically significant differences in mean scores for the nine Conceptions of Assessment scales for each amount of assessment literacy training (Table 36). Thus, like assessment type, the amount of assessment training this group of teachers has had made no difference to their conceptions of assessment.

Table 36. Multivariate Results COA-III Subscales by Assessment Literacy Training

Effect	*n*	Hypothesis *df*	Wilks' λ	*F*	*p*
No Training	76	9	.98	1.38	.20
Some Preservice Hours	183	9	.99	.48	.89
Workshop or Seminar	220	9	.98	1.18	.31
Completed Undergraduate Paper	114	9	.97	1.57	.12
Completed Postgraduate Paper	28	9	.99	.48	.89

Conclusion

The meaning of teachers' conceptions of assessment is understood by looking at the structural relationships of the various conceptions in the model and the differing levels of agreement or support that teachers have for each conception. The measurement model permits such an analysis. The model proposed has four major conceptions (i.e., student accountability, school accountability, improvement, and irrelevance) with different internal structures. The first two conceptions are first-order factors, while the latter two are second-order factors constituted of three to four first-order factors. The intercorrelation of the four main factors provides the greatest insight into teachers' conceptions of assessment.

It is noted that the relationship between improvement and irrelevance is inverse. That is, if teachers think assessment is about improvement then it is unlikely they will consider assessment as irrelevant ($r = -.69$) and they are likely to believe that assessment is connected to accountability of schools or teachers ($r = .58$). This unexpected relationship may be because of the impact of self-management of New Zealand schools wherein teachers are accountable for the effectiveness of their work in changing student learning outcomes to their colleagues and to a school-based Board of Trustees made up of parents of pupils. Teachers who conceive of assessment as

improvement tended to have just moderate likelihood of agreeing that assessment is about certifying student performance or achievement. This may be because of the impact of student-centred conceptions.

If teachers think assessment is about school accountability, then they may or may not believe that assessment is irrelevant; belief in one is independent of belief in the other. Teachers who believe in assessment as school accountability are highly likely to also conceive assessment as for the student accountability and improvement. This suggests a nexus of conceptions around the idea that assessment for school accountability may lead to a raising of educational standards that will in turn lead to improved ability of students to receive qualifications and recognition of achievement. This is what some advocates of high-stakes accountability testing have argued would and should happen (e.g., Resnick & Resnick, 1989). However, it is worth nothing that this effect is found in a context where there is no externally mandated national test, just a program of school-based policies on assessment for school-based management and information.

Finally, when teachers think assessment is about student accountability, it is moderately likely they will also consider assessment to be irrelevant, because it is bad for students or inaccurate, such that they can safely ignore it. It is possible that this conception is related to strong student centred learning beliefs or humanistic curriculum or nurturing teaching beliefs. Teachers who conceive of assessment as student accountability are likely to have only a weak relationship to improvement. In other words, assessment of students is likely to be irrelevant when it is connected to accountability but is more likely to be acceptable if it is related to improvement of teaching and learning.

The pattern found among New Zealand teachers of positive attitudes toward improvement conceptions and more negative attitudes towards accountability conceptions is consistent with, though more complex than, results from the limited number of studies on teachers' conceptions of assessment conducted elsewhere in the world. Philippou and Christou (1997) found, in terms of the mathematics curriculum, that Greek and Cypriot teachers strongly agreed with using assessment for improvement (i.e., diagnosing students' difficulties, and evaluating the effectiveness of instruction), but were less supportive of assessment for accountability (i.e., assigning grades to students) and disagreed with assessment having a role in modifying the centrally determined curriculum. Warren and Nisbet (1999, p. 517), in a study of Australian teachers' uses of assessment, found that "primary teachers seemed to use assessment more often to inform the teacher with regard to teaching than to inform the learner with regard to learning and that using assessment for reporting to others was not as important as informing teaching and learning". Saltzgaver (1983) found, when describing the dominant conceptions of assessment of just one Australian teacher, ten convictions that could be mapped onto two of the major assessment conceptions found in this research (i.e., improvement and irrelevance). Stamp (1987) identified three major conceptions of assessment among pre-service teacher trainees in Australia; specifically an improvement type conception of identifying individual student learning needs for the purpose of catering for those individual requirements; an irrelevance-type conception requires teachers to conduct assessment even though they get in the way of students' creativity and intuition as much as their academic development; and a student accountability-type conception that uses tests and examinations to collect end-of-course information about students partly in order to motivate them to compete for more marks.

There is partial support for the models advanced by Gipps et al. (1995) and M. Hill (2000b). Statements based on the various types of assessment use (e.g., head note, systematic, evidence gathering, and systematic planners) tended to appear in the theoretically appropriate areas (i.e., irrelevance, student accountability, and improvement respectively). Although, the data analysed here do replicate the uses of assessment found in the Gipps et al. and Hill findings, they do not support the same tripartite structure. The structure of the conceptions of assessment model developed in this research is significantly fuller or more complete than the descriptions of three archetypes of assessment use in that it shows how teachers' conceptions of assessment are structured and interrelated.

There is clear indication from the distribution of teachers' levels of agreement (Table 27) that not all conceptions are held equally strongly by all teachers. Yet, the generalisability of the model was tested by examining three major questions: (a) whether teachers' conceptions are stable across all population characteristics, (b) whether teachers' conceptions are stable across the types of assessments teachers use and their assessment training, and (c) whether differing amounts of assessment training shape conceptions of assessment. The data showed that mean scores did not differ by teachers' different population characteristics, definition and practice of assessment or assessment literacy training.

Most importantly, it is necessary to examine how teachers' conceptions of assessment relate to their instructionally related conceptions of learning, teaching, curriculum, and efficacy. Study 1 hinted that there was a significant impact of assessment on those related practices and conceptions. However, it was not possible to establish the nature of those relationships other than by the interpretive insight of the researcher. In order to address the issue of how teachers' conceptions of assessment

relate to these constructs, it is necessary to collect simultaneous data. Such an approach was taken in Study 6 where teachers supplied data about their conceptions of assessment, teaching, learning, curriculum, and efficacy.

This chapter has documented the successful development through three studies of a 50 item self-report instrument to measure teachers' multi-dimensional conceptions of assessment. Teachers' conceptions of assessment related to four main ideas; school accountability, student accountability, improvement of teaching and learning, and irrelevance of assessment. These conceptions involved agreement towards the correlated conceptions of assessment is improvement of teaching and learning and assessment measures school accountability and a rejection of the correlated conceptions of assessment is student accountability and assessment is irrelevant. The model of teachers' conceptions of assessment established in this research, while consistent with other research and common sense, is more sophisticated and complex than existing descriptions of how teachers conceive of assessment. The model of teachers' conceptions of assessment and its instrument are a significant contribution to research into this area.

CHAPTER IV. A FOUR-FACET MODEL OF TEACHERS' INSTRUCTIONAL CONCEPTIONS

Delandshere and Jones (1999) proposed that there are two major foci of teachers' conceptions to do with learning, curriculum, and assessment. The first, associated with accountability, is a subject-centred approach that emphasises teachers' transmission of rules and facts assessed for sanction and verification of whether or not the student has learned the content. In contrast, the second, associated with improvement, is a learner-centred approach that emphasises students' construction of knowledge through learning experiences assessed for the formative purpose of documenting learning and providing feedback. Although, this conceptualisation of teachers thinking echoes the narrow view of "formative assessment good, summative assessment bad" discussed in the earlier chapter on conceptions of assessment, it may be widely held by teachers (see Scriven, 1991 for a detailed critique of the use and misuse of these terms).

However, it is important to test whether such a pattern of conceptions is adequate to explain teachers' conceptions. To begin to test the nature of teachers' conceptions about learning, curriculum, teaching, assessment, and efficacy, it is necessary to collect data about teachers' conceptions in all five areas. Study 6 reports and analyses data collected about teachers' conceptions of assessment, as discussed in the previous chapter as Study 5, and relates that to their conceptions of learning, curriculum, teaching, and their own efficacy.

In addition, Study 6 tests the generalisability of the CoA-III model across participant population characteristics and across assessment uses, practices, and training. It is important to know whether school or teacher characteristics have

significant impact on teachers' conceptions and whether teachers' use of assessment tools, their assessment practices, or their assessment training affect the type of assessment conceptions they have. Evidence for the stability of these conceptions would suggest that teachers' conceptions vary only according on idiosyncratic rather than predictable bases.

Data were collected through self-report Likert-type response questionnaires using previously published instruments. Teachers' conceptions were analysed and the resulting exploratory pattern was tested in a measurement model using SEM. That process showed how teachers' various conceptions and demographic characteristics related to each other.

Study 6: CoA-III Related to Teachers' Conceptions of Learning, Curriculum, Teaching, and Teacher Efficacy

As discussed in Study 5, this study comprised two survey questionnaires that explored teachers' responses to a series of statements about assessment, assessment types, assessment practices, teaching, curriculum, learning, and teacher efficacy. It is predicated on the model of teachers' conceptions of assessment developed in Study 5 and pre-existing instruments to measure five conceptual areas. This chapter documents the instruments used to collect data about the various conceptions, reports the measurement model structure of each instrument, its correlations with the CoA-III and the other measured conceptions, and finally, describes how the five key conceptions interrelate in one structural model.

Instruments

A number of other instruments were used to elicit information from teachers about their conceptions of learning, curriculum, teaching, and teacher efficacy. In

151

addition, an assessment practices instrument was used to identify the frequency of

teacher use of a variety of assessments. These instruments are described next.

Conceptions of Learning

As used in the earlier study on teacher and student conceptions about studying,

six items from the Entwistle (n.d.) *Approaches and Study Skills Inventory for Students*

(Assist) were used to measure teacher conceptions about learning. This instrument has

six statements that elicit responses along the Marton and Saljo (1976) taxonomy of

learning conceptions. Three statements are designed to probe surface conceptions of

learning (i.e., getting facts and information, remember things well, and using the

information I've got); while three statements probe deep conceptions of learning (i.e.,

developing as a person, seeing things in a different and more meaningful way, and

understanding new material for myself). It should be noted that the statement about

applying or using information could be interpreted either as a deep or surface approach

and that Study 1 had dropped this item from analysis but it did not clearly load on either

concept. It was anticipated that the larger sample size of Study 6 would clarify the

status of this item.

Conceptions of Curriculum

Cheung's (2000) conceptions of curriculum inventory was adapted to New

Zealand circumstances by making small wording changes. For example, the item about

consummatory experience, a term introduced by Eisner and Vallance (1974), was

rewritten "Curriculum should try to provide satisfactory consumer experience for each

student". The 21 items grouped into four major perspectives (i.e., academic,

humanistic, technological, and social reconstruction) are listed with their loadings on

the appropriate factor (Table 37). Note that the statements all have strong loadings on

their respective factors and that the whole inventory had good fit to the model in

Cheung's (2000) research with teachers.

Table 37. Conceptions of Curriculum Inventory Statements, Factors, and Loadings

Statements	Loading
Academic Subjects	
The basic goal of curriculum should be the development of cognitive skills that can be applied to learning virtually anything.	.72
School curriculum should aim at developing students' rational thinking.	.59
Curriculum should require teachers to transmit the best and the most important subject contents to students.	.54
School curriculum should aim at allowing students to acquire the most important products of humanity's intelligence.	.54
Curriculum should stress refinement of intellectual abilities.	.50
Humanistic	
Students' interests and needs should be the organising centre of the curriculum.	.64
Curriculum and instruction are actually inseparable and the major task of a teacher is to design a rich learning environment.	.62
The ultimate goal of school curriculum should help students to achieve self-actualisation.	.62
Curriculum should try to provide satisfactory consumer experience for each student.	.56
Teachers should select curriculum contents based on students' interests and needs.	.54
Technological	
Curriculum and instruction should focus on finding efficient means to a set of predetermined learning objectives.	.68
Curriculum should be concerned with the technology by which knowledge is communicated.	.65
Learning should occur in certain systematic ways.	.60
I believe that educational technology can increase the effectiveness of students' learning.	.59
Sections of curriculum content and teaching activities should be based on the learning objectives of a particular subject.	.57
The learning objectives of every lesson should be specific and unambiguous.	.50
Social Reconstruction	
Existing problems in our society should be organising centre of curriculum.	.80
Curriculum should let students understand societal problems and take action to establish a new society.	.75
Curriculum contents should focus on societal problems such as pollution, population explosion, energy shortage, racial discrimination, corruption, and crime.	.67
The most important goal of school curriculum is to foster students' ability to critically analyse societal problems.	.60

Conceptions of Teaching

For brevity's sake, given the large number of responses required by participants, it was decided to select three statements for each of the five perspectives in the *Teaching Perspectives Inventory* (Pratt & Collins, 1998). Teacher responses to the TPI have been collected from a number of cross-cultural studies and collected into a database of over 1,000 respondents. From that dataset, the three strongest loading statements, based on equamax rotation factor analysis, for each subscale were identified (J. B. Collins, personal communication, August 23, 2001). Statements covering the three aspects of each perspective (i.e., a belief, an intention, and an action) were selected. The statements selected for this study with their perspective, aspect, and equamax factor loading are identified in Table 38. Loadings are all acceptably high but it should be noted that this instrument has not yet been tested with SEM that may identify potential psychometric improvements.

Table 38. Teaching Perspectives Inventory Factors, Statements and Loadings

Factors and Statements	Equamax Factor Loadings
Apprenticeship	
I link the subject matter with real settings of practice or application	.59
My intent is to demonstrate how to perform or work in real situations	.69
To be an effective teacher, one must be an effective practitioner	.53
Development	
I challenge familiar ways of understanding the subject matter	.59
My intent is to help people develop more complex ways of reasoning	.67
Teaching should focus on developing qualitative changes in thinking	.57
Nurturing	
I encourage expressions of feeling and emotion	.73
My intent is to build people's self-confidence and self-esteem as learners	.77
In my teaching, building self-confidence in learners is priority	.73
Social Reform	
I help people see the need for changes in society	.78
I expect people to be committed to changing our society	.81
Individual learning without social change is not enough	.66
Transmission	
I make it very clear to people what they are to learn	.55
My intent is to prepare people for examinations	.63
Effective teachers must first be experts in their own subject areas	.52

Conceptions of Teacher Efficacy

The 16-item *Teacher Efficacy Scale* (Gibson & Demo, 1984) was composed of two largely uncorrelated ($r = -.19$) subscales (i.e., personal teaching efficacy had 9 items and general teaching efficacy had 7 items). Guskey and Passaro (1994) revised this instrument and found two (similarly uncorrelated $r = -.24$), factors that they identified as Internal and External dimensions of teacher efficacy. These located teaching effectiveness either in the teacher's personal ability or in external environmental factors such as home or family environment.

Because a significant number of the items in the Guskey and Passaro revision were very similar in wording the ten most strongly loading items that provided maximally unique statements about each scale were taken from the Guskey and Passaro (1994) teacher efficacy instrument. This reduction in the number of items was also

156

required to limit the total number of responses required by each participant. The scale

statements adopted for this study, with their loadings, are listed in Table 39. Loadings

are acceptable but it should be noted that more sophisticated SEM analysis has not yet

been conducted with these statements.

Table 39. Teacher Efficacy Statements, Factors, and Loadings

Factors and Statements	Varimax Loadings
Internal	
If a student masters a new concept quickly, this might be because the teacher knew the necessary steps in teaching that concept.	.62
When a student gets a better grade than he/she usually gets, it is usually because I found better ways of teaching that student.	.60
When a student does better than usually, many times it is because the teacher exerts a little extra effort	.55
When I really try, I can get through to most difficult students.	.53
If a student in my class becomes disruptive and noisy, I feel assured that I know some techniques to redirect him/her quickly.	.44
External	
I am very limited in what I can achieve because a student's home environment is a large influence on his/her achievement.	.78
Teachers are not a very powerful influence on student achievement when all factors are considered.	.66
The hours in my class have little influence on students compared to the influence of their home environment.	.56
I have not been trained to deal with many of the learning problems my students have.	.45
When a student is having difficulty with an assignment, I often have trouble adjusting it to his/her level.	.42

Results

The first step was to ascertain the properties of each conception scale in order to

establish whether the data collected fit the model characteristics of each instrument.

Data were tested, where applicable, according to various authors' design using SEM

measurement modelling. Having established that each instrument provided satisfactory

fit to the data, subscale mean scores and standard deviations were calculated. The

resulting 22 subscales (nine from the CoA-III and 13 from the related concepts) were

correlated. As would be expected with sample sizes exceeding 200, many of the

correlations between the various subscales were statistically significant. Interpretation has focused on larger correlations (i.e., those greater than .30) and those where no significant correlation were found. Because two different forms were used the sample size was reduced by about half for analysis of the five conceptions data.

The correlation pattern suggested that there were some underlying meta-conceptions or conceptions that integrated the 22 conceptions subscales. An exploratory factor analysis was undertaken, which in turn, suggested a meaningful structure to teachers' conceptions of assessment, teaching, curriculum, learning, and teacher efficacy. That model was tested using SEM.

Conceptions of Learning

Like Study 1, documented earlier, only two conceptions of learning were detected; that is deep and surface. The deep factor contained four statements that focused on understanding, transforming, developing personally, and applying, while the surface factor contained only two statements that focused on memory and facts or details. Note that unlike Study 1, the use or application of information statement did load on the deep learning conception, as was predicted by the surface—deep learning model. This difference may point to the differing conceptions of learning between the secondary school teachers in Study 1 and the primary school teachers in Study 6.

Intercorrelation of the two factors was moderate ($r = .393$, $p<.01$) while fit to the model (Figure 10) was excellent ($\chi^2 = 10.593$; $df = 19$; $TLI = .999$; $RMSEA = .025$). Teachers agreed strongly with the deep conception of learning ($N = 234$, $M = 5.15$, $SD = .61$) and only moderately with the surface ($N = 236$, $M = 3.85$, $SD = .90$) conception.

158

Figure 10. Conceptions of Learning Measurement Model
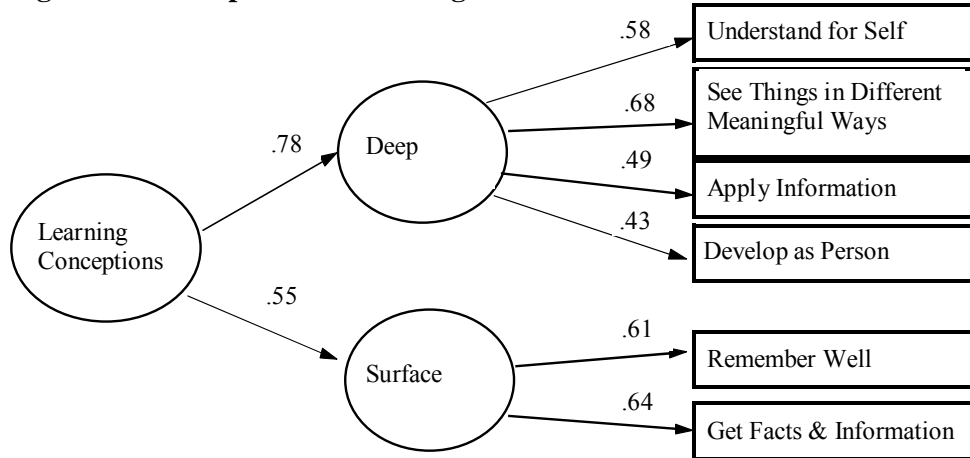
The deep conception of learning correlated moderately with only three of the improvement conceptions (i.e., describe student ability, improve student learning, and improve teaching) (Table 40). The surface conception of learning, however, correlated moderately to strongly with all of the improvement conceptions and the two accountability conceptions. Further, the surface conception of learning also correlated significantly but weakly with the inaccuracy of assessment conception.

Table 40. Intercorrelations Conceptions of Learning and CoA-III

	Conception of Learning	
Conceptions of Assessment	Deep	Surface
Accountability School System	.07	.31**
Accountability Students Learning	.10	.41**
Improve Describe Students	.30**	.33**
Improve Student Learning	.25**	.23**
Improve Teaching	.32**	.25**
Improve Valid	.05	.28**
Irrelevance Bad	-.13	.02
Irrelevance Ignore	-.08	-.02
Irrelevance Inaccurate	.12	.14*

Note. *p<.05; **p<.01

Thus, the deep conception of learning appeared to be somewhat connected to the improvement of teaching and nothing else. The surface conception, although also

connected to the improvement assessment conceptions (*r* between .23 and .33), was

most strongly connected to assessment for student accountability (*r* = .41). It may be

that the relatively lower level of agreement given to the surface conception of learning

may be connected to its strong association with the student accountability conception

which itself received low levels of agreement.

The deep conception of learning had statistically significant correlations with all

related conceptions except for the social reconstruction curriculum conception and

external teacher efficacy subscales, while the surface conception had statistically

significant correlations with all the related conceptions except for the external teacher

efficacy scale (Table 41).

Table 41. Intercorrelations Conceptions of Learning and Related Conceptions

	Conception of Learning	
Related Conceptions	Deep	Surface
Curriculum		
Technological	.23**	0.31**
Academic	.33**	0.36**
Humanistic	.38**	0.30**
Social Reconstruction	.03	0.18**
Teaching		
Apprenticeship	.38**	0.27**
Cognitive Development	.43**	0.20**
Nurturing	.58**	0.25**
Social Reform	.22**	0.23**
Transmission	.14*	0.36**
Teacher Efficacy		
External	-.10	0.04
Internal	.18**	0.28**

Note. **p<.05; **p<.01*

The deep conception correlated most strongly with the nurturing perspective of

teaching (*r* = .58) and the humanistic conception of curriculum (*r* = .38). In contrast,

the surface conception correlated most strongly with the academic conception of

curriculum (*r* = .36) and the transmission conception of teaching (*r* = .36). The

moderate intercorrelation of internal teacher efficacy and surface learning (*r* = .28)

suggested that teachers believe that they can improve student learning at the surface level.

By examining the strongest correlations, two clearer conceptions may be identified. The deep conception grouped with the cognitive development and nurturing conceptions suggesting that teachers believe students need careful assistance into deep cognitive development. The surface conception grouped with academic conceptions of curriculum and transmission approaches to teaching suggesting that the kinds of assessments used to make students accountable are oriented to getting facts and knowledge and that teachers disagree with this approach to learning and assessment.

Conceptions of Curriculum

The four-factor structure of Cheung's Conceptions of Curriculum instrument was evaluated using SEM (Figure 11). Loadings of items on their respective factors, and of the factors onto the model were robust though overall fit exceeds the threshold of *RMSEA* ≤ .08 (χ^2 = 599.593; *df* = 183; *TLI* = .965, *RMSEA* = .098).

Note that this measurement model is only a first-order correlated model in contrast to Cheung's hierarchical model. Note also that the academic and humanistic conceptions are, at least in the responses of this group of New Zealand teachers, to all extents identical conceptions. Clearly, this group of teachers conceived that humanistic conceptions are nearly identical to academic conceptions; in other words improving students' lives involves improving their academic and intellectual abilities as well. This suggests that future research should be conducted with a three-factor model.

Figure 11. Conceptions of Curriculum Measurement Model
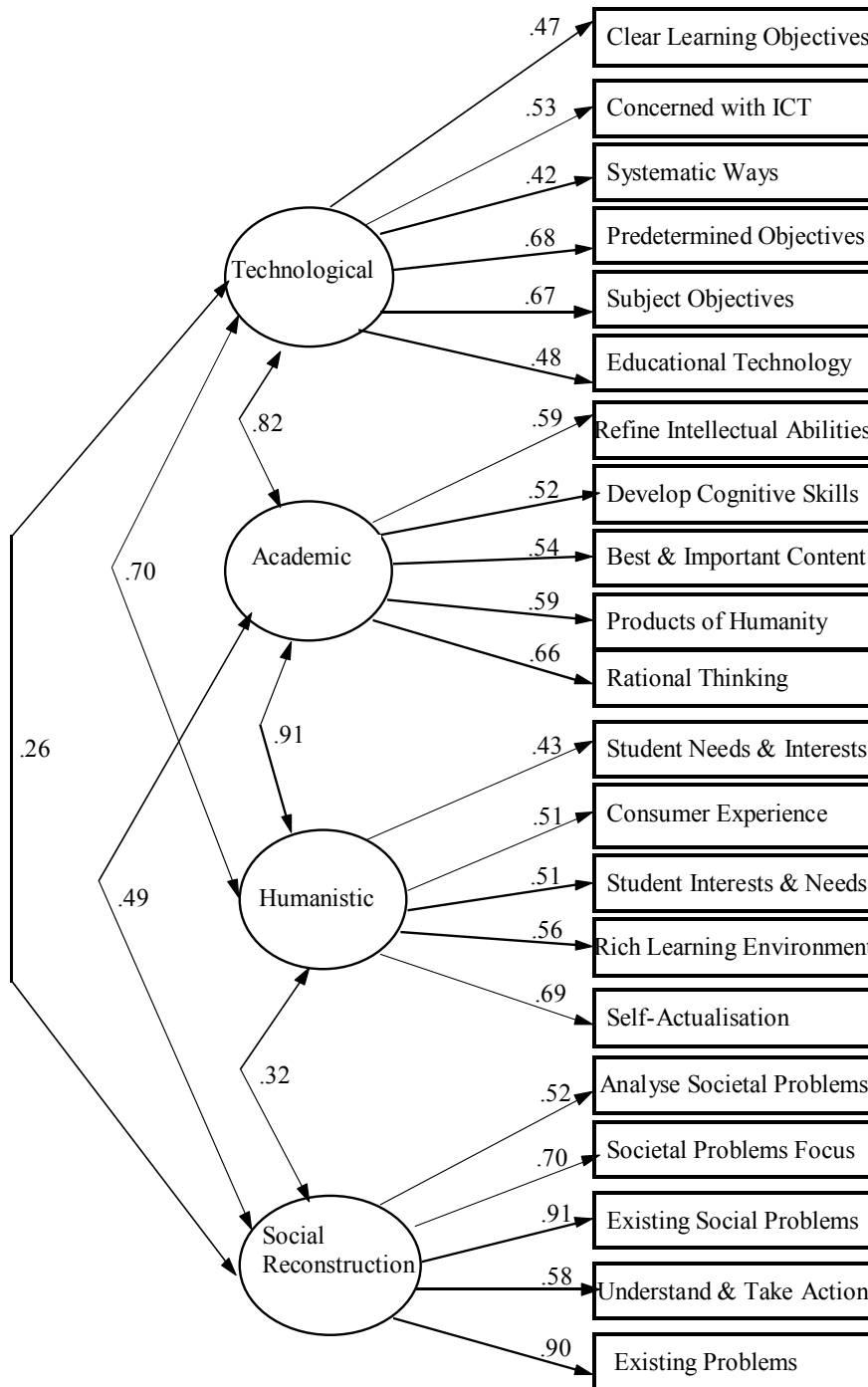
Scale intercorrelations were moderate to very strong (Table 42) with the academic conception being highly correlated with the humanistic ($r = .91$) and the technological ($r = .82$) conceptions. Teachers tended to agree strongly with the humanistic conception ($N = 217$, $M = 4.68$, $SD = .74$), while they moderately agreed with both the

academic ($N = 211$, $M = 4.27$, $SD = .74$) and technological ($N = 227$, $M = 4.29$, $SD = .70$) conceptions. They only slightly agreed with the social reconstruction conception ($N = 225$, $M = 3.02$, $SD = .85$).

Table 42. Intercorrelations Conceptions of Curriculum Subscales

Subscale	1	2	3	4
1. Technological	—			
2. Academic	0.82	—		
3. Humanistic	0.70	0.91	—	
4. Social Reconstruction	0.26	0.49	0.32	—

Three of the curriculum conceptions, excluding the social reconstruction conception, loaded moderately on both accountability and the four improvement subscales (Table 43). The social reconstruction conception also correlated moderately on the two accountability conceptions, the improvement of student learning and validity conceptions and weakly on the bad and inaccurate irrelevance conception subscales. The irrelevance assessment subscales had zero correlations with the curriculum conceptions except for the inaccurate subscale that correlated weakly with the academic, humanistic, and social reconstruction conceptions.

Table 43. Intercorrelations Assessment Practices and CoA-III

Conceptions of Assessment	Conception of Curriculum			
	Technological	Academic	Humanistic	Social Reconstruction
Accountability School System	.37**	.42**	.27**	.34**
Accountability Students Learning	.41**	.40**	.29**	.32**
Improve Describe Students	.48**	.43**	.29**	.11
Improve Student Learning	.40**	.37**	.35**	.20**
Improve Teaching	.38**	.35**	.36**	.07
Improve Valid	.37**	.31**	.22**	.22**
Irrelevance Bad	-.11	-.02	-.02	.14*
Irrelevance Ignore	-.15*	-.05	-.04	.07
Irrelevance Inaccurate	.08	.23**	.18**	.19**

Note. *$p<.05$; **$p<.01$

All four conceptions of curriculum subscales correlated moderately with the five teaching perspectives subscales, both the conceptions of learning subscales, and the

internal teacher efficacy subscale (Table 44). None of the curriculum conceptions

subscales correlated significantly with the external teacher efficacy subscale. The one

exception to this pattern, as noted before, was seen in the zero correlation of the deep

learning subscale with the social reconstruction curriculum conception subscale.

Table 44. Intercorrelations Assessment Practices and Related Conceptions

| Related Conceptions | Conceptions of Curriculum | | | |
	Technological	Academic	Humanistic	Social Reconstruction
Teaching Perspectives				
Apprenticeship	.38**	.41**	.42**	.27**
Cognitive Development	.36**	.45**	.32**	.32**
Nurturing	.24**	.32**	.50**	.14*
Social Reform	.21**	.31**	.23**	.54**
Transmission	.35**	.37**	.20**	.42**
Learning				
Deep	.23**	.33**	.38**	.033
Surface	.31**	.36**	.30**	.18**
Teacher Efficacy				
External	.05	.03	-.02	.07
Internal	.23**	.39**	.24**	.32**

Note. * p<.05; **p<.01

By focusing on only the strongest correlations (i.e., $r > .40$) between the

curriculum conceptions and the CoA-III and the related instructional conceptions, it is

possible to identify four groups of conceptions. One is a grouping of social

reconstruction, social reform, and transmission conceptions (a conception of telling for

societal change); a second groups together the humanistic, apprenticeship, and nurturing

conceptions (a conception focused on student centred learning); a third brings together

the academic, apprenticeship, cognitive development, school and student accountability

conceptions (a conception that involves intellectual development assessed for

accountability), and the fourth relates the technological, school accountability, describe

students, and improve student learning conceptions (a conception that focuses on

systematic, data collection for improvement and accountability). Further analysis is

needed to determine whether these four groups of conceptions exist.

Conceptions of Teaching

The structure of the TPI was tested in an SEM measurement model and overall

fit was good (χ^2 = 277.062; df = 80; TLI = .975; $RMSEA$ =.069) (Figure 12). The five

factors had moderate to strong intercorrelations (Table 45). As might be expected from

the literature on teacher-oriented versus student-oriented conceptions of teaching, the

nurturing perspective had low correlations with social reform and transmission

perspectives. Note that this measurement model is only a first-order correlated model

in contrast to the proposed hierarchical model. Note also that the apprenticeship and

cognitive development conceptions are, at least in the responses of this group of New

Zealand teachers, to all extents highly similar conceptions. This correlation is similar to

the pattern of correlations between the humanistic and academic conceptions of

curriculum. In other words, the student-oriented approach to teaching that involves

modelling required behaviour is highly related to developing students' cognitive

abilities as well. This suggests that future research should be conducted with a four-

factor model. Nevertheless, the integrity of the factors was sufficient for use in this

study.

Table 45. Intercorrelations Teaching Perspectives Subscales

Subscales	1	2	3	4	5
1. Apprenticeship	—				
2. Cognitive Development	.87	—			
3. Nurturing	.71	.64	—		
4. Social Reform	.52	.76	.32	—	
5. Transmission	.48	.72	.20	.65	—

Teachers agreed strongly with the Nurturing (N = 237, M = 5.43, SD = .56),

Apprenticeship (N = 234, M = 4.92, SD = .66), and Developmental (N = 232, M = 4.70,

SD = .74) perspectives. Only a moderate level of agreement was given to the Social

Reconstruction (*N* = 234, *M* = 3.87, *SD* = 1.03) perspective while slight agreement was

given to the Transmission (*N* = 235, *M* = 3.37, *SD* = .78) perspective. These results are

consistent with Pratt and Collins' (2001) reported results for teachers-in-training of

whom nearly 70% had nurturing perspective as dominant and only six percent had

transmission as their dominant perspective.

Figure 12. Conceptions of Teaching Measurement Model
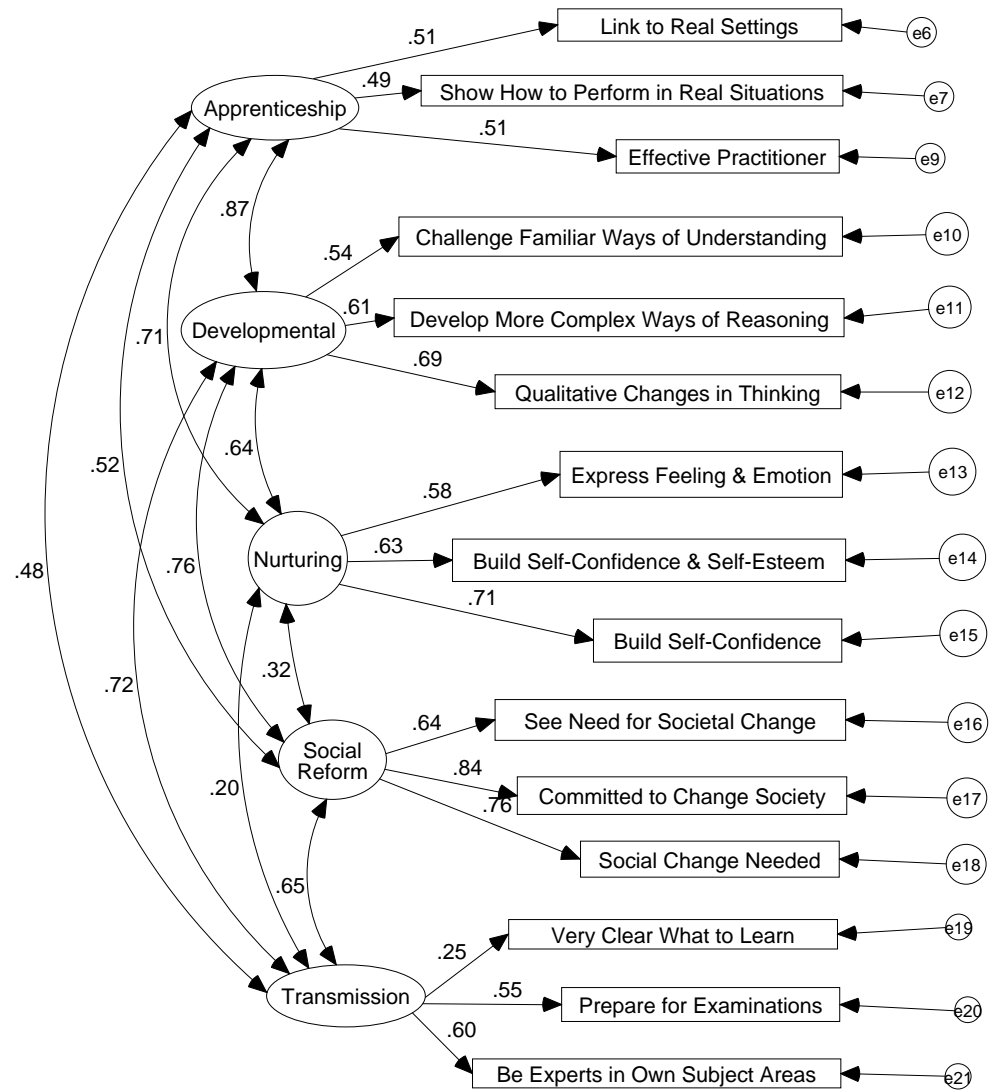
Not surprisingly, only the nurturing perspective did not correlate with the two accountability subscales with values not significantly different from zero (Table 46). In contrast, the transmission perspective correlated positively and strongly on the two accountability subscales. In only two cases did the five teaching perspectives not correlate positively and significantly on the four improvement subscales. Specifically, social reform did not correlate on the improvement of teaching subscale, and the

nurturing perspective did not correlate with the validity subscale. In only three cases

did any of the teaching perspectives correlate positively and significantly with any of

the irrelevance subscales. Both the nurturing and social reform perspectives correlated

weakly with the inaccurate assessment subscale, while the social reform subscale

likewise correlated weakly with the ignore assessment subscale.

Table 46. Intercorrelations of Teaching Perspectives and CoA-III

Conceptions of Assessment	Conception of Teaching				
	Apprenticeship	Cognitive Development	Nurturing	Social Reform	Transmission
Accountability School System	.29**	.26**	.12	.31**	.40**
Accountability Students Learning	.22**	.24**	.08	.20**	.45**
Improve Describe Students	.27**	.26**	.22**	.16*	.30**
Improve Student Learning	.35**	.28**	.24**	.19**	.23**
Improve Teaching	.31**	.23**	.23**	.10	.14*
Improve Valid	.13*	.20**	.09	.22**	.35**
Irrelevance Bad	-.08	-.06	-.07	.08	.09
Irrelevance Ignore	-.09	.01	-.05	.16*	.06
Irrelevance Inaccurate	.10	.11	.13*	.15*	-.05

Note. * *p*<.05; ***p*<.01

This pattern of correlations suggested that teachers with a social reform

perspective tended to not see a connection between assessment and the improvement of

teaching or that assessment was bad for students. They did however tend to conceive of

assessment as something to be ignored and as something inherently inaccurate while at

the same time they predominantly conceived of assessment as a tool of accountability.

Nurturing teachers connected assessment to the description of student learning and to

the improvement of both teaching and learning, while having a tendency to conceive

assessment as inaccurate. Teachers with a transmission perspective strongly associated

assessment with valid, descriptions of assessment learning for the purpose of student

and school accountability. Both the apprenticeship and development perspectives are

associated with accountability and improvement conceptions of assessment while being zero related to the irrelevance conceptions.

Except for the external teacher efficacy subscale, all correlations between the teaching perspectives and related conceptions are positive, statistically significant, and weak to moderate (Table 47). The apprenticeship perspective correlated most strongly with the academic, technological, and humanistic curriculum, the deep learning, and the internal teacher efficacy subscales. The cognitive development perspective associated most strongly with the academic curriculum and deep learning subscales. The nurturing perspective correlated most strongly with the humanistic curriculum and deep learning perspectives. The social reform perspective correlated most strongly with the social reconstruction curriculum subscale. The transmission perspective correlated most strongly with the social reconstruction curriculum, followed next by the technological and academic curriculum and surface learning subscales.

Table 47. Intercorrelations Teaching Perspectives and Related Conceptions

Related Conceptions	Teaching Perspective				
	Apprenticeship	Cognitive Development	Nurturing	Social Reform	Transmission
Curriculum					
Technological	.38**	.36**	.24**	.21**	.35**
Academic	.41**	.45**	.32**	.31**	.37**
Humanistic	.42**	.32**	.50**	.23**	.20**
Social Reconstruction	.27**	.32**	.14*	.54**	.42**
Learning					
Deep	.38**	.43**	.58**	.22**	.14*
Surface	.27**	.20**	.25**	.23**	.36**
Teacher Efficacy					
External	-.03	-.11	-.18**	.03	.08
Internal	.33**	.17**	.14*	.27**	.36**

Note. * p<.05; **p<.01

By taking into account only those correlations >.40 it is possible to infer that teachers' conceptions relate to two main ideas; that is improving students' learning and

transmission-accountability. Teachers with an apprenticeship perspective tended to also

believe in assessment for improvement of learning and teaching, and in academic and

humanistic conceptions of curriculum. The nurturing perspective is likewise related to

the humanistic conception of curriculum and also to deep views of learning. The

cognitive perspective related strongly to the academic conception of curriculum and

also to deep view of learning. Thus one conception might be that the teachers

conceived of teaching as a process of improving the deep capabilities of students' minds

in a student-friendly way. On the other hand, the transmission view tended to correlate

strongly with social reconstruction view of curriculum and the school and student

accountability conceptions of assessment. So it is possible teachers think of teaching as

being accountable for changing society.

Conceptions of Teacher Efficacy

Like both the previously referred to teacher efficacy instruments (Gibson & Dembo,

1984; Guskey & Passaro, 1994) the intercorrelation of the two factors was basically

zero ($r = -.004$, $p=.953$). This lack of meaningful relationship was reflected in the low

loadings of the two factors on the Teacher Efficacy construct (internal $\lambda = .17$; external

$\lambda = .14$) (Figure 13). Nevertheless, the quality of the measurement was satisfactory

because the overall fit of the model was good ($\chi^2 = 131.742$; $df = 31$; $RMSEA = .074$;

$TLI = .972$) and the reliability of subscales was adequate (internal $\alpha = .65$; external

$\alpha = .65$). In addition, the loadings of the statements on the two factors ranged from

acceptable ($\lambda = .25$) to strong ($\lambda = .76$). Teachers gave moderate agreement ($N = 236$,

$M = 3.82$, $SD = .62$) to their personal internal ability to effect student learning outcomes

and slight agreement to the power of external ($N = 237$, $M = 2.86$, $SD = .90$) factors to

interfere with their ability to effect learning outcomes. Thus, on the whole this group of

teachers believed that they were personally able to effect changes in student learning

and that external factors did not really impinge on their efficaciousness.

Figure 13. Conceptions of Teacher Efficacy Measurement Model
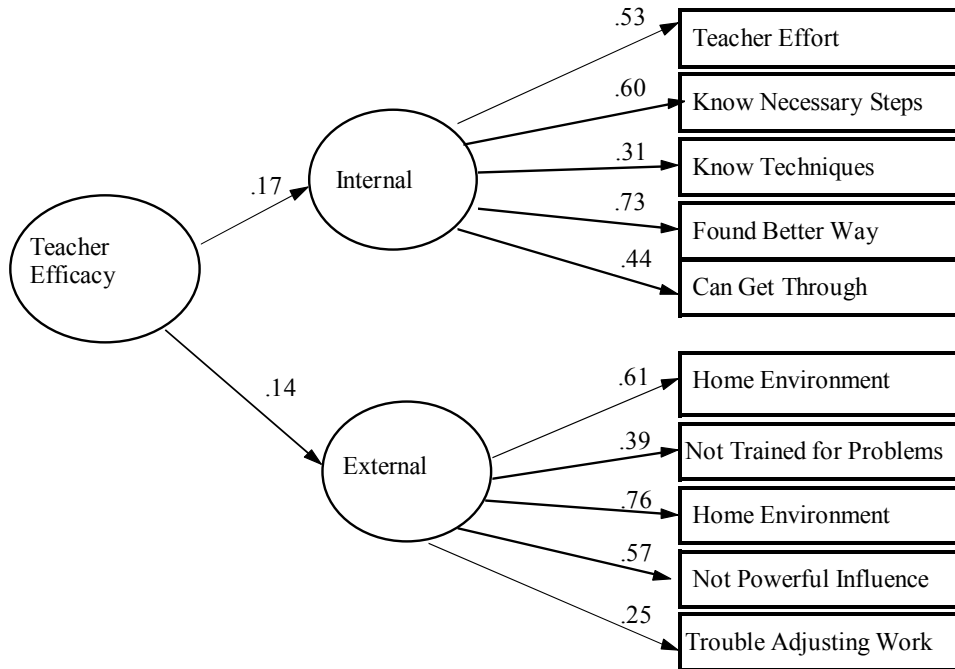
Higher scores on external factors on teacher efficacy were weakly correlated

with increases in conception of assessment as a negative force (Table 48). This is

shown in the positive correlations with the conception of assessment as something bad

or to be ignored, the negative, though weak, correlations with improvement of teaching

and learning subscales, and the weak, positive correlation with the use of assessment for

student accountability. On the other hand the internal efficacy was moderately

correlated with the two accountability conceptions and weakly correlated with all four

of the improvement conceptions. It had zero correlation with the three irrelevance

scales.

Table 48. Intercorrelations Conceptions of Efficacy and CoA-III

	Conception of Teacher Efficacy	
Conceptions of Assessment	External	Internal
Accountability School System	.04	.39**
Accountability Students Learning	.19**	.30**
Improve Describe Students	-.08	.30**
Improve Student Learning	-.17**	.24**
Improve Teaching	-.17**	.21**
Improve Valid	-.06	.27**
Irrelevance Bad	.30**	.09
Irrelevance Ignore	.25**	.08
Irrelevance Inaccurate	.09	.08

Note. * p<.05; **p<.01

The external teacher efficacy scale had zero correlations with all of the other conceptions except for a weak, negative correlation with the nurturing perspective of teaching (Table 49). In contrast, the internal teacher efficacy scale had statistically significant correlations with all other conceptions, though strongest correlations ($r > .30$) were seen with the academic and social reconstruction conceptions of curriculum, and the apprenticeship and transmission conceptions of teaching.

Table 49. Intercorrelations Conceptions of Efficacy and Related Conceptions

	Teacher Efficacy	
Related Conceptions	External	Internal
Curriculum		
Technological	.05	.23**
Academic	.03	.39**
Humanistic	-.02	.24**
Social Reconstruction	.07	.32**
Teaching Perspectives		
Apprenticeship	-.03	.33**
Cognitive Development	-.11	.17**
Nurturing	-.18**	.14*
Social Reform	.03	.27**
Transmission	.08	.36**
Learning		
Deep	-.10	.18**
Surface	.04	.28**

Note. * p<.05; **p<.01

This group of teachers believed themselves to be internally efficacious and that belief correlated, surprisingly, most of all with accountability conceptions rather than improvement conceptions. This suggested that teachers believe themselves capable of meeting external accountability requirements, whether those are of themselves or of their teachers. Furthermore, this internal efficacy related to a traditional academic, transmission oriented conception of teachers' work; in other words, teachers believe they can transmit to students the material they need for academic development.

Teachers' Instructional Conceptions

The relationship of the COA-III to each of teachers' separately analysed conceptions about teaching, learning, curriculum, and teacher efficacy left a somewhat muddled and inadequate understanding of how the various conceptions inter-relate. It is more useful to analyse how those conceptions relate in one structured analysis. To this end, exploratory factor analysis using maximum likelihood extraction method with oblimin rotation (Table 50) was conducted. Each of the 22 scale scores was treated as an observed variable, instead of the latent factor it actually is, because of the low ratio of cases to variables.

Four meaningful meta-factors, each containing between four and seven scale scores, were found and upon inspection of constituent variables were named; external checking, influencing learning, student centred learning, and telling for change. External checking loaded strongly on the two accountability conceptions of assessment, the technological curriculum conception, and the surface conception of learning. The two teacher efficacy factors loaded weakly on this factor, indicating that teachers' two differing confidence beliefs (i.e., internal "I can do it" and external "Outside school causes mean I can't do it") relate to the surface learning dominated assessment used to

make schools, teachers, and students accountable. This is somewhat unexpected, as it suggests that teachers' do not have confidence that they are able to effect change in student learning outcomes in terms of either a student-centred learning, or improvement assessment, or transmission for change constituted conceptual patterns. The influence learning factor loaded strongly on the four improvement conceptions of assessment and equally strongly but inversely on the three irrelevance conceptions of assessment. The student centred learning factor loaded strongly on two of the curriculum conceptions (i.e., academic and humanistic), two teaching perspectives (i.e., nurturing, and apprenticeship) and on the deep conception of learning. Note that this factor loaded on none of the assessment conceptions. The telling for change factor loaded on the three teaching perspectives of social reform, transmission, and cognitive development and on the social reconstruction conception of curriculum. Note again that this factor loaded on none of the conceptions of assessment.

Table 50. Factor Structure Teachers' Instructional Conceptions

Factors	Factor Loadings			
	1	**2**	**3**	**4**
External Checking				
Accountability Students Certification	**0.67**	-0.09	0.01	0.12
Accountability School System	**0.54**	0.26	-0.06	0.23
Curriculum: Technological	**0.48**	0.18	0.35	0.01
Learning: Surface	**0.46**	0.02	0.10	0.08
Efficacy: Internal	**0.38**	0.01	0.08	0.19
Efficacy: External	**0.32**	-0.30	-0.13	-0.05
Influence Learning				
Irrelevance Ignore	0.06	**-0.71**	-0.05	0.18
Irrelevance Bad	0.23	**-0.69**	-0.04	0.05
Improve Student Learning	0.23	**0.68**	0.15	0.10
Improve Describe Students	0.44	**0.62**	0.09	0.04
Improve Teaching	0.21	**0.60**	0.31	-0.08
Improve Valid	0.42	**0.59**	-0.20	0.20
Irrelevance Inaccurate	0.08	**-0.50**	0.41	-0.05
Student Centred Learning				
Curriculum: Humanistic	0.31	0.02	**0.64**	-0.04
Teaching: Nurturing	-0.15	0.05	**0.61**	0.22
Learning: Deep	-0.05	0.09	**0.56**	0.13
Curriculum: Academic	0.47	0.03	**0.48**	0.10
Teaching: Apprenticeship	0.04	0.08	**0.44**	0.32
Telling for Change				
Teaching: Social Reform	-0.04	-0.09	0.03	**0.77**
Teaching: Cognitive Development	-0.14	0.04	0.31	**0.68**
Teaching: Transmission	0.30	0.01	-0.11	**0.60**
Curriculum: Social Reconstruction	0.27	-0.14	0.05	**0.45**

The intercorrelations between the four teacher conceptions factors was low ranging from close to zero to a maximum of $r = -.38$ (Table 51). Thus, the four factors are largely independent of each other, though there was a weak inverse relationship between external checking and telling for change and a weak positive correlation between student-centred learning and telling for change.

Table 51. Intercorrelations Teachers' Instructional Conceptions Factors

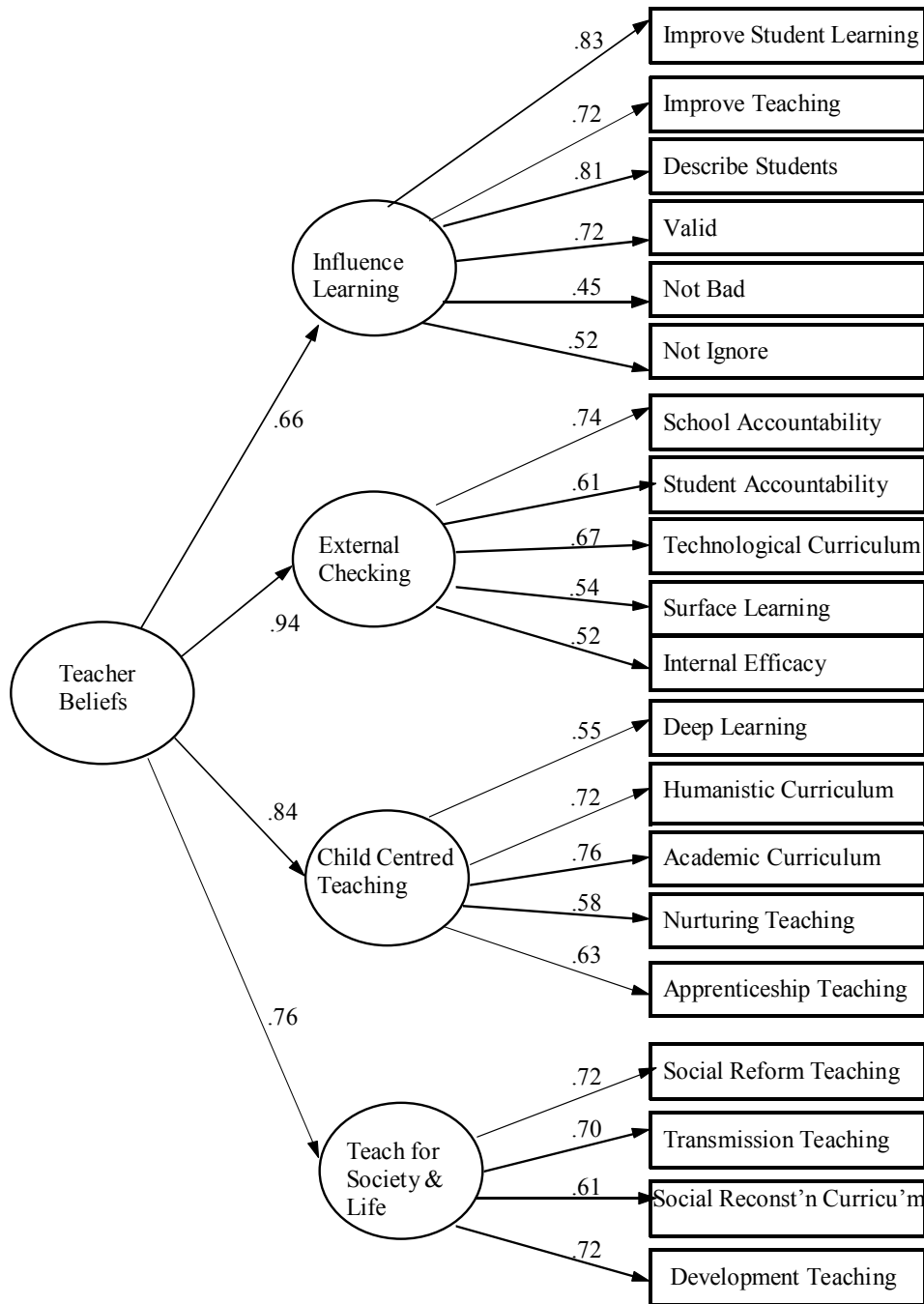| Factor	Factors			
	1	2	3	4
1. External Checking	—			
2. Influence Learning	-0.08	—		
3. Student Centred Learning	-0.18	0.16	—	
4. Telling for Change	-0.38	0.11	0.34	—

For ease of interpretation, the irrelevance scales were reverse scored so that they would load in the same direction as the other improvement scales. This resulted in seven scale scores making up the influence-learning factor that showed stronger levels of agreement as scores increased. Note that through the planned missing data collection matrix only half of the teachers completed the related conceptions questionnaire. Thus, there was a reduced sample size for any factor not totally dependent on COA-III scales only (see Table 52 for number of valid cases for each factor).

Table 52. Descriptive Statistics Teachers' Instructional Conceptions Factor Scores

Factor	Descriptive Statistics			Effect Size			
	N	*M*	*SD*	1	2	3	4
1. External Checking	222	3.28	.48	—			
2. Influence Learning	525	4.06	.58	1.47	—		
3. Student Centred Learning	195	3.21	.34	.17	1.75	—	
4. Telling for Change	214	2.12	.39	2.67	4.00	2.99	—

The four factors solution of teacher conceptions was tested with SEM (Figure 14). One of the scale scores (i.e., external teacher efficacy) loaded very weakly ($\lambda = .05$) and was removed from the structural equation. The resulting fit of the measurement model to the data was good ($\chi^2 = 823.262$; *df* = 166; *RMSEA* = .086; *TLI* = .968). The external checking and student centred learning factors were the strongest loading factors ($\lambda = .94$ and $\lambda = .84$ respectively), while the telling for change and assessment influences learning factors loaded acceptably ($\lambda = .76$ and $\lambda = .66$ respectively).

Figure 14. Four Factor Teachers' Instructional Conceptions Measurement Model

The strength of agreement of teachers to each factor was calculated by

converting scale scores to the agreement scale values ranging between 1 and 6 (Table

54). Teachers moderately agreed with the assessment influences learning factor,

177

slightly agreed with the external checking and student centred learning factors, and disagreed with the tell for change factors. Effect sizes, calculated for the differences in the mean factor scores between the various factors using sample size weighted pooled variances (Table 52), indicated that all differences were large except for that between student-centred learning and external checking. Teachers agreed that assessment was something that influenced their teaching and student learning and which could improve both. Teachers agreed, albeit less strongly, that assessment makes schools, teachers, and students accountable. That conception, however, is associated strongly with conceptions that learning is surface, that curriculum is a technological means to an end, and that teachers are capable of achieving student learning outcomes. This was interpreted to suggest that the assessment used in accountability measures surface learning of facts and information and that teachers accept it as a means of accountability because surface learning and its assessment are things that teachers believe they are able to do through the application of a systematic technological approach. Perhaps, this is a reflection of the 'tick, cross, slash' approach described as so prolific among New Zealand primary school teachers documented by M. Hill (2000b).

Teachers also slightly agreed with student focused conceptions of humanistic and academic conceptions of curriculum, nurturing and apprenticeship perspectives of teaching, and deep views of learning. This factor associated a deep view of learning (e.g., learning something for oneself, or understanding things in a new way) with a caring approach to helping students develop as whole people. It is worth nothing that the student-centred philosophy here is not anti-intellectual in its understanding of what student development means, since it is associated with an academic approach to curriculum that stresses (a) refinement of intellectual abilities, (b) the development of cognitive skills that can be applied to learning virtually anything, (c) the transmission of

the best and the most important subject content, (d) students' acquiring the most important products of humanity's intelligence, and (e) developing students' rational thinking. This factor was independent of any of the assessment conceptions found in the external checking and influences learning and teaching factors. Perhaps this was so because true student centred learning and teaching is so intangible and wrapped up in the subjectivity of teacher-student relationships and modelling that it cannot be measured or assessed at all.

The fourth branch to this model of teachers' conceptions was one that they disagreed with. It was expected that primary teachers would disagree with the transmission or banking (Barnes, 1976) type of teaching. It is also not unexpected that teachers would not conceive of primary schooling as a method of reconstructing or reforming society. What is surprising is the association of an intellectual development perspective on teaching being associated negatively with transmission and social reconstruction. It was anticipated that the cognitive development statements (i.e., challenging familiar ways of understanding subject matter, helping people develop more complex ways of reasoning, and developing qualitative changes in thinking) would associate with the deep learning conception. It can only be assumed that to teachers these statements seemed too much like the negatively perceived social reform or reconstruction statements (e.g., helping people see the need for changes in society, fostering students' ability to critically analyse societal problems, understanding societal problems and taking action to establish a new society). This socially conservative position is reminiscent of claims that schools are agents of social reproduction not transformation (Bourdieu, 1974; Harker, 1982).

The structure of this measurement model has another striking characteristic, that is, the separation of assessment for external checking versus influence learning

purposes. This dichotomy reflects other research, among which are Carr's (2001) accountability-oriented folk model of assessment versus an improvement-oriented alternative model, Torrance and Pryor's (1998) accountability-oriented convergent assessment contrasted with teaching-improvement or divergent model of assessment, and Philipp, Flores, Sowder, and Schappelle's (1994) evaluation for reporting contrasted with assessment used to inform teaching. These models are themselves reflections of the discredited false dichotomy between 'summative bad' and 'formative good' models of conceiving assessment. Most importantly, this model is not framed around a simple dichotomy; it is multi-dimensional. Simple opposites do not explain how teachers conceive of assessment or, for that matter, how teachers believe how assessment, teaching, learning, and assessment meld into the professional practice of instruction.

Two-fold models of how teachers' conceptions of assessment, learning, teaching, and curriculum are not uncommon (e.g., Delandshere & Jones, 1999; Kember & Kwan, 2000). Those models tend to propose a negative content-centred (material oriented, transmissive teaching, summative assessment) contrasted with a learning centred (student oriented, facilitative teaching, formative assessment) approach to teaching. This four-facet model is an advance on those two-fold models of teachers' conceptions of learning, teaching, curriculum, and assessment. This data supports a four-facet view that (a) assessment improves the quality of teaching and learning, (b) external checking is something that teachers associate with their own efficaciousness, perhaps because it measures surface learning, (c) student-centred deep learning is not associated directly with assessment, and (d) the transmission view of teaching does not associate with the accountability view, but rather with a changing society and cognitive development approach.

There is greater similarity to the model of professional competence in Dwyer and Villegas' (1993) description of four broad domains of teacher life. The domain of teaching for student learning is quite similar to the assessment influences learning factor while the domain of creating an environment for student learning is significantly more like the student-centred learning factor. Less similar, though potentially linked, are the domains of teacher professionalism, perhaps related to the external accountability factor and the organising content knowledge for student learning which may be equivalent to the transmission teaching dimension in the telling for change factor.

CHAPTER V. CONCLUSION

The purpose of this research was to investigate teachers' thought processes relevant to learning and assessment. The research explored the nature of teachers' conceptions of assessment, ascertained the structure of those conceptions, and determined how those various conceptions of assessment related to each other, and examined how teachers' conceptions of assessment related to teachers' conceptions about learning, teaching, curriculum, and teacher efficacy. Furthermore, it examined how teachers' conceptions of assessment related to teacher and school characteristics, and to teachers' assessment practices (i.e., importance and type of assessments used). In addition, it determined the strength of agreement that teachers had for each conception. This chapter summarises the methods used and the findings, discusses implications of the findings, and outlines future research possible as a consequence of this research.

A series of research questions were addressed using a combination of qualitative and quantitative methods, with an emphasis on advanced correlational techniques, to develop and test models of how teachers' conceptions were structured. Six consecutive studies were conducted into teachers' conceptions of learning, assessment, and related practices and conceptions. Findings were related to the research questions and, because sample sizes in Studies 5 and 6 were large and representative enough, the findings were generalisable to the population of New Zealand primary school teachers.

Specifically, the following questions were addressed. Studies 1 and 2 addressed the conceptions of assessment and learning that New Zealand teachers have and whether those conceptions could be mapped to the categories found in the literature. Studies 3 to 5 answered questions related to the structure of each major purpose-defined

conception of assessment (e.g., is it multi-faceted), while Study 5 addressed the question of how teachers' conceptions related to each other. Study 6 addressed several important questions including; (i) What assessment methods do teachers associate with the term 'assessment' and what relationship is there between those association and conceptions of assessment?, (ii) What assessment practices do teachers use and what relationship is there between uses and teachers' conceptions of assessment?, (iii) What relationship is there between individual teacher characteristics (i.e., gender, role, experience, assessment training) and teachers' conceptions of assessment?, (iv) What relationship is there between school characteristics (e.g., size, SES) and teachers' conceptions of assessment?, (v) Are the models proposed for the teachers' conceptions instruments (i.e., learning, curriculum, teaching, and teacher efficacy) reflected in New Zealand teachers' responses?, (vi) How do teachers' conceptions of assessment relate to their conceptions of teaching, learning, curriculum, and teacher efficacy?, and (vii) Is there a meaningful structure that relates teachers' conceptions of assessment, curriculum, teaching, learning, and efficacy?

Data were collected for the most part through anonymous survey questionnaires designed to reflect teachers' conceptions about the various constructs. The conceptions were expressed as statements about major conceptions derived from the literature and from qualitative data collected from teachers, which had been collected through free response statements and interviews. Data were collected from samples of practising secondary and primary teachers, teacher trainees, and undergraduate education students. Standard procedures for developing self-report attitude questionnaires were used involving iterative analyses to determine dependable instruments of teachers' conceptions and to eliminate non-fitting attitude statements (Gable & Wolf, 1993).

Self-report Likert-type data were analysed initially with exploratory factor analysis, using maximum likelihood estimation with oblimin rotation whenever sample sizes permitted. Factors were inspected to ensure that statements within each factor were theoretically meaningful and that the items making up each factor had sufficient psychometric characteristics (i.e., loading >.30, loadings on one unique factor, three or more statements per factor). Once exploratory factor analysis established a possible measurement model, structural equation modelling was undertaken to determine the fit of the data to the proposed measurement model. Inspection of fit indices, error variances, and modification indices indicated whether the model had sufficient psychometric characteristics.

Findings

A brief overview in response to the research questions is listed here, before a wider discussion of the findings is undertaken. It was found that the theoretical models of learning (surface-deep) and assessment (student accountability, school accountability improvement, irrelevance) could be used to describe the views New Zealand teachers had about learning and assessment. A 50-statement, self-report attitude inventory about teachers' conceptions of assessment was developed that mapped teachers' conceptions of assessment into a multi-faceted and hierarchical model. Specifically, nine conceptions were found, seven of which mapped on to two second-order factors that contained multiple first-order factors (i.e., improvement and irrelevance). The conceptions of assessment model had four major inter-correlated facets (i.e., improvement, irrelevance, school accountability, student accountability). Three of the conceptions were positively intercorrelated, whereas the fourth major conception (i.e., irrelevance) was negatively correlated with improvement, uncorrelated with school

accountability, and positively correlated only with the student accountability conception. Teachers agreed with the improvement and the school accountability conceptions, gave slight agreement to the student accountability conceptions, and disagreed with the irrelevance conception. In addition, it appeared that the model of teachers' conceptions of assessment was robust across multiple populations made up of primary level teacher-trainees, undergraduate education students, and practising primary school teachers. Thus, a complex model of teachers' conceptions of assessment found support in the responses of New Zealand primary school teachers.

Eleven types, forms or methods of assessment reduced to four categories (i.e., teacher controlled classroom assessments, formal examinations, oral assessments, and portfolio) and these categories did not provide statistically significant different mean scores for the nine teachers' conceptions of assessment. Three major categories of assessment practices were found (i.e., deep cognitive processing, informal classroom assessment, and formal assessment) and these, also, did not produce statistically significant different mean scores for the nine teachers' conceptions of assessment. Except for one of the nine conceptions and only for teacher role, individual teacher characteristics (i.e., gender, role, experience, assessment training) likewise did not generate statistically significant different mean scores among the nine teachers' conceptions of assessment. It was also found that school characteristics (e.g., size, SES) provided statistically similar mean scores for the nine teachers' conceptions of assessment. Views of secondary and primary teachers in this research were similar; likewise views of in-service teachers and pre-service teacher trainees and undergraduate education students were similar. Thus, teachers' conceptions of assessment were remarkably universal and stable across population characteristics and self-reported assessment practices.

Data from New Zealand teachers' responses to inventories of conceptions (i.e., learning, curriculum, teaching, and teacher efficacy) were found to conform to the proposed models. A surface and deep factor for learning conceptions and four conceptions of curriculum (i.e., academic, technological, social reconstruction, and humanistic) were found. Two teacher efficacy factors (i.e., internal and external) and five perspectives of teaching were observed (i.e., social reform, transmission, nurturing, apprenticeship, and cognitive development). These instruments provided data that were consistent with the theoretical models on which they were based. A meaningful structure that relates teachers' conceptions of assessment, curriculum, teaching, learning, and efficacy was found. That model proposed four major groupings of teachers' conceptions (i.e., external checking, assessment influence learning, student-centred learning, and telling for change) that were largely uncorrelated. Teachers were found to disagree with the telling for change conception, have weak agreement with the student-centred learning and the external checking conceptions, and moderate agreement with the assessment influences learning conception.

As suggested by Clark and Peterson (1986), teachers' conceptions of assessment have complex structures and relationships. Teachers' instructional conceptions were modelled as four major conceptions; assessment that influenced learning, student centred teaching, external checking of teachers and students, and telling for change.

Thus, an efficient instrument has been developed that measures teachers' conceptions of assessment regardless of teacher background variables or assessment practices and that can be used in conjunction with other instruments and measures to identify the structure of teachers' conceptual processes relevant to the art and act of instruction.

Implications

A number of implications can be drawn from this research related to self-regulation theory, design of teacher education, and assessment policy implementation. Self-regulation theory (Zimmerman, 2001) argues that strategy control depends on declarative knowledge of multiple strategies, procedural ability to operate various strategies, and meta-cognitive control, awareness, and monitoring of strategy use and effectiveness. The model of teachers' conceptions of assessment developed here suggests that having differing conceptions may be appropriate contingent on varying conditions of accountability and improvement in the educational setting. It may be useful for teachers to be able to choose or emphasise a different conception of assessment depending on factors influencing teacher work. In order to select another conception of assessment it would be necessary, according to self-regulation theory, for teachers to have knowledge about the different conceptions, their own conceptions, and the appropriate grounds for different conceptions. With this type of knowledge, and meta-cognitive control teachers may be better position to achieve desired educational outcomes.

Further, this model of teachers' conceptions of assessment could be put to use in teacher professional development and policy contexts. The implementation of any new assessment policy, tool, or practice, whether at the national or local school level, needs to take account of the complex structure of teachers' conceptions of assessment to ensure success. Kahn (2000) pointed out that teachers appeared to assimilate new assessment practices (e.g., constructivist, deep) into long-standing transmission, teacher-oriented, accountability type assessment and learning frameworks. Certainly, the implementation of new standards from professional bodies or state authorities, while

well intentioned, may be reduced in effectiveness if teachers' conceptions of assessment remain unchanged or unchallenged, or if teachers remain unaware of their own conceptions. Likewise, teacher professional pre-service preparation and in-service development in the area of assessment needs to take account of teachers' pre-existing conceptions, if it is to be effective in moving teachers toward a preferred structure of conceptions.

As a case in point, there was a marked difference in emphasis on differing conceptions of assessment between the principal and several of the teachers in one school that participated in CoA-II. In that study, the principal agreed strongly with the improvement conception and disagreed with the accountability conception. In contrast, three of the teachers had much higher agreement on irrelevance and accountability conceptions and disagreed with the improvement conception. Fundamentally, despite talking about the common word 'assessment' these teachers were talking past each other. A new improvement-oriented assessment policy or practice in that school, without explicit attention to the inappropriate accountability and irrelevance conceptions of the teachers, would likely be adopted and assimilated into a traditional model of assessment as something to be used but ignored.

In terms of professional development of teachers' conceptualisation of assessment, there is a need to make explicit the different understandings teachers may have of assessment to ensure that participants do not talk past each other. The instrument developed in this thesis may be useful in disrupting the tendency to simply adopt and adapt rather than actually change practices because it could be used to make more explicit the conceptions teachers have about assessment and trigger such discussions. It was found during the development of the instrument, for example, from the results of ten teachers in the same school, that individuals had markedly differing

profiles in their conceptions of assessment. This suggested that making explicit and resolving differences would enhance the success of any innovation around 'assessment' and certainly this would need to be resolved before the introduction of any improvement-oriented assessment policy or practice. Similarly, pre-service instruction in assessment should make explicit varying conceptions of assessment, their rationales and consequences, and attempt to move future teachers and managers away from a simplistic dichotomy of formative good—summative bad.

There is also an opportunity to utilise the structure of teachers' conceptions of assessment in policy contexts. This research showed that teachers agree with the conception that assessment improves teaching and learning and reject assessment's irrelevance in this context. Further, this positive attitude to improvement is simultaneously paralleled with much less support for making students accountable and by an association of external checking assessment with surface learning. This mix suggests that the introduction of assessment innovations should be done in such a way as to minimise association with external accountability dimensions (whether those be at the school, teacher, or student levels) and instead maximise association with teachers' individual capability to improve their own instruction and the learning of their own students—at least if the aim is to improve student learning outcomes. Thus, in the context of the hypothetical conversation used, in the introduction to illustrate the range of teachers' conceptions, it would behove those introducing assessment changes in New Zealand to structure not only the rhetoric, but also the implementation of any new assessment tool as something to be used by classroom teachers for improvement and not as something used by school managers and leaders for accountability purposes. In other jurisdictions, where large-scale externally mandated assessment is the norm, teachers may have more pronounced irrelevance and student accountability conceptions that

would need to be addressed should policy require a greater shift towards improvement oriented conceptions.

The structure of teachers' conceptions also suggests that externally mandated assessments are perceived as associated with surface learning and that those assessments are divorced from student-centred development, learning, or growth. This suggests that the legally mandated introduction of national assessments could result in teachers being able to ensure most students achieve but that teachers would continue to believe those assessments were divorced from the core of classroom life. New Zealand already has externally mandated checking of student performance, required by national educational goals and administrative guidelines, against curriculum objectives and levels. This has led teachers to develop accountability conformity systems and procedures that promote their continually generating accountability data rather than the use of assessment to raise standards and improve student achievement (M. Hill, 2000b). The research reported here reinforces the conclusion that externally mandated assessments could be implemented but, should such assessments be implemented, teachers would not believe that the assessments would relate to the improvement of learning and teaching and development of the whole student. Disassociating assessment systems from accountability purposes may be the approach most likely to be effective in raising standards and performance.

Because New Zealand has no externally mandated assessments for primary school students in place, there exists the possibility of improving teachers' assessment literacy through the implementation of an assessment innovation. By helping teachers implement an improvement practice of assessment and by associating that with the deep learning, student-centred philosophy and by making teachers accountable for the process within their own institutions rather than to some outside agency, it may be

possible to connect with teachers' dominant conceptions effectively. Emphasis on a school-based, and managed process of improvement-oriented evaluation of student assessment results, is likely to result in educational improvement in the quality of teaching and the quality of student learning outcomes (see for example the SEMO model, Timperley & Robinson, 2002). The implication of this research is that the focus in assessment policy should not be on compulsion but rather on identifying and responding to teachers' conceptions because no matter what policies are put in place unless teachers' conceptions are addressed the policy change will be ineffective.

A further implication of this research is the feasibility of using the COA-III instrument in research contexts. In any research conducted to evaluate the effectiveness of interventions designed to change or modify teachers' conceptions of assessment the instrument could be used to monitor changes in teachers' conceptions. Additionally, the instrument could be used to investigate the origins of conceptions of assessment.

Future Research

The results of this research, although robust, require follow-up. It is important to examine the stability of the COA-III model across ethnicity, school sectors, and school roles. To achieve this focused studies are required, specifically with teachers of different ethnic backgrounds (e.g., Maori, Pacific Nation, or Asian), with principals, and teachers in intermediate and secondary schools. This would test the generalisability of the model, which has been based on a relatively homogeneous population, with more heterogeneous populations of teachers. By implication, at least 500 teachers in each category would be required to test the stability across new populations.

Another research question that merits investigation is to do with how the instructional beliefs model of Study 6 performs if all the observed items had been used

instead of the 22 latent variables. Sample size of several thousand is needed to permit specification of a model that utilises all the observed and latent variables.

Additionally, it would be useful to identify a meaningful measure of teacher effectiveness so that a structural equation model of how teachers' conceptions relate to student learning outcomes could be developed and tested. The data here suggest that teachers with an influence learning perspective may be more focused on deep student learning than those who emphasise external checking. The data also suggest that external checking emphasis would predict students who demonstrate surface rather than deep learning outcomes. The method of measuring the quality of student outcomes on the surface-deep axis was used in identifying effective teachers (Bond, Smith, Baker, & Hattie, 2000) suggesting that rather than focus simply on total scores or grades that the architecture of student learning could be used as a relatively objective measure of teacher effectiveness. Besides naturalistic data collection, it may be possible to investigate teachers and students with matching and mismatched conceptions of assessment in the manner of aptitude-treatment interaction studies. With such a measure and sample size, a full SEM analysis of teachers' conceptions could be conducted and any difference in conceptions between expert, competent, and novice teachers could be detected.

Furthermore, having proposed that the CoA-III inventory could be used in professional development it would be worthwhile to determine whether the interpretations from use of the instrument contributes to changing teachers' conceptions of assessment. More importantly, it would be useful to attempt to determine if there is a 'best' conception of assessment profile that produces the greatest improvement in student learning and teacher instruction.

Further research into the relationship of teachers' conceptions and their actual practices is warranted. It would be useful to determine if teachers' conceptions can be discovered in their actual practice. Research elsewhere has used a variety of methods (e.g., case studies, naturalistic class observations, interviews with teachers and students, narrative descriptions of own work, stimulated recall from class videotapes, student essays, faculty reflection on own case descriptions, ratings by students) to investigate such relationships (Hativa, 2000). Evidence for linkages between teaching conceptions (i.e., traditional, transmission oriented, behaviourism vs. constructivist) and learning of teacher trainees was obtained at the lesson level using methods such as concept maps, post mapping questionnaire, stimulated recall interviews, and short answer assessment (Jensen, Kauchak, & Rowley, 2001). Thus, a wide variety of qualitative approaches exist that would help determine the consistency of self-professed conceptions obtained through survey questionnaire with teachers' practices or conceptions as expressed through action or intention. Part of the point in investigating the relationship of practice to conception is to answer, in addition to the consistency issue, questions such as: Do accountability teachers behave differently to improvement teachers; do conceptions predict assessment practice; and do assessment conceptions relate systematically to teaching or curriculum practices?

Having begun this research with an interest in both teachers' and students' conceptions of learning, it is only appropriate that investigation of students' conceptions of assessment be conducted. Work would have to be done to adapt the present teacher instrument to the linguistic and cognitive capabilities of students, before investigation into the relationship of teachers' and students' mutual and possibly interactive conceptions could begin. Research into students' conceptions of assessment may add light on the origins of teachers' conceptions of assessment, since Pajares (1992) has

argued that teachers' conceptions area product of their educational experiences as students.

Contribution

This thesis research has aimed to contribute to the fields of teachers' conceptions and educational assessment. It has created a relatively brief self-report instrument that aims to make teachers' conceptions of assessment explicit and which could be used in professional development (e.g., make teachers think about and possibly change their conceptions) within schools for clarification of conceptions among faculty especially between leaders and teachers, and the development of assessment innovations, especially informing implementation and design.

More importantly, the thesis has identified some unexpected structures and relationships among teachers' conceptions. Among these findings are the existence of the irrelevance conception, the positive connection of school accountability and improvement conceptions, the positive relationship of teacher efficacy to external checking accountability assessment, the splitting of deep and surface views of learning with the former unconnected to assessment and the latter connected to external accountability checking assessment, and the positive relationship of irrelevance only with student accountability. It also discovered, again surprisingly, that the structure of teachers' conceptions was universal and stable across all population characteristics (at least among those participating in the various studies in this thesis).

Finally, this thesis' greatest contribution is the confirmation that teachers' conceptions of assessment are complex, hierarchical, multidimensional, and interrelated, not simple or dichotomous. Researchers, policy makers, teacher educators, teachers, and teacher trainees now have evidence from teachers' own thinking that assessment should not be conceived in a superficial simplistic fashion. Continuing to

portray assessment as a bipolar construct along a positive--negative dimension is to

perpetuate a misconception of assessment to the detriment of learning and teaching as

well as the quality of assessment.

REFERENCES

Adams, T. L., & Hsu, J.-W. Y. (1998). Classroom assessment: Teachers' conceptions and practices in mathematics. *School Science and Mathematics, 98*(4), 174-180.

Airasian, P. W. (1997). *Classroom assessment*. (3rd ed.). New York: McGraw-Hill.

American Federation of Teachers (AFT), National Council for Measurement in Education (NCME), & National Education Association (NEA), (1990). Standards for teacher competence in educational assessment of students. *Educational Measurement: Issues and Practice, 9*(4), 30-32.

Anthony, G. (1994). Learning strategies in the mathematics classroom: What can we learn from stimulated recall interviews? *New Zealand Journal of Educational Studies, 29*(2), 127–140.

Anthony, G. (1997). Task demands and instructional support in mathematics: Influences on strategic learning behaviours of secondary students. *set: research information for teachers,* (2), 6.

Arbuckle, J. L. (1999). AMOS (Version 4.01). [Computer software]. Chicago: SmallWaters Corp.

Asch, R. L. (1976). Teaching beliefs & evaluation. *Art Education, 29*(6), 18-22.

Bandura, A. (1989). Human agency in social cognitive theory. *American Psychologist, 44*(9), 1175-1184.

Barnes, D. (1976). *From communication to curriculum.* London: Penguin Press.

Bempechat, J., & Drago-Severson, E. (1999). Cross-national differences in academic achievement: Beyond etic conceptions of children's understandings. *Review of Educational Research, 69*(3), 287–314.

Biggs, J. B. (1987a). *Student approaches to learning and studying*. Research

    monograph. Melbourne, Australia: ACER.

Biggs, J. B. (1987b). *Learning Process Questionnaire: Manual*. Melbourne, Australia:

    ACER.

Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO*

    taxonomy (Structure of the Observed Learning Outcome). New York: Academic

    Press.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Educational*

    Assessment: Principles, Policy and Practice, 5(1), 7-74.

Bond, L., Smith, T., Baker, W. K., & Hattie, J. A. (2000). *The certification system of the*

    National Board for Professional Teaching Standards: A construct and

    consequential validity study. (Research Report). Greensboro, NC: University of

    North Carolina, Center for Educational Research and Evaluation.

Borko, H., Flory, M., & Cumbo, K. (1993). *Teachers' ideas and practices about*

    assessment and instruction: A case study of the effects of alternative assessment

    in instruction, student learning and accountability practices. (CSE Tech. Rep.

    366). Los Angeles: Graduate School of Education, National Center for Research

    on Evaluation, Standards, and Student Testing.

Borko, H., Mayfield, V., Marion, S., Flexer, R., & Cumbo, K. (1997). Teachers'

    developing ideas and practices about mathematics performance assessment:

    Successes, stumbling blocks, and implications for professional development.

    Teaching & Teacher Education, 13(3), 259-278.

Bourdieu, Pierre (1974). The school as a conservative force: scholastic and cultural

    inequalities. In J. Eggleston (Ed.) *Contemporary Research in the Sociology of*

    Education, (pp. 32-46). London: Methuen..

Broadfoot, P. (2002). *Learnacy or lunacy? Assessing the 21st century curriculum.* Paper

    presented to the International Association for Educational Assessment (IAEA)

    Annual Conference, September 1 – September 6, 2002, Hong Kong.

Brown, G. T. L. (2002). Student Beliefs about Learning: New Zealand Students in Year

    11. *Academic Exchange Quarterly, 6*(1), 110-114.

Brown, G. T. L. (in review). Student use of agreement and frequency formats in self-

    reports: Which works better? *Psychological Reports*

Browne, M. W., & Cudek, R. (1989). Single sample cross-validation indices for

    covariance structures. *Multivariate Behavioral Research, 24,* 445-455.

Browne, M. W., & Cudek, R. (1993). Alternative ways of assessing model fit. In K. A.

    Bollen & J. S. Long (Eds.), *Testing structural equation models,* (pp. 136-162).

    Newbury Park, CA: Sage.

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A

    theoretical synthesis. *Review of Educational Research, 65*(3), 245–281.

Butterfield, S., Williams, A., & Marr, A. (1999). Talking about assessment:

    mentor-student dialogues about pupil assessment in initial teacher training.

    Assessment in Education, 6(2), 225–246.

Calderhead, J. (1996). Teachers: Beliefs and knowledge. In D. C. Berliner & R. C.

    Calfee (Eds.), *Handbook of educational psychology* (pp. 709-725). New York:

    Simon & Schuster Macmillan.

Campbell, J., Brownlee, J., & Smith, D. (1996). The differential impact of teachers'

    approaches to teaching on secondary students' approaches to learning. *Education

    Research and Perspectives, 23*(2), 95–111.

Cannell, J. J. (1989). *How public educators cheat on standardized achievement tests.*

    Albuquerque, NM: Friends for Education.

Carr, M. (2001). *Assessment in early childhood settings : learning stories*. London :
Paul Chapman.

Chang, E., & Simpson, D. (1997). The circle of learning: Individual and group
processes. *Education Policy Analysis Archives, 5*(7). [On-line] Retrieved
September 16, 1998, from http://olam.ed.asu.edu/epaa/v5n7/

Cheung, D. (2000). Measuring teachers' meta-orientations to curriculum: Application of
hierarchical confirmatory analysis. *Journal of Experimental Education, 68*(2),
149-165.

Cheung, D., & Ng, P.-H. (2000). Science teachers' beliefs about curriculum design.
Research in Science Education, 30(4), 357-375.

Chou, C.-P., & Bentler, P. M. (1995). Estimates and tests in structural equation
modeling. In R. H. Hoyle, (Ed.). *Stuructural equation modeling: Concepts,
issues, and applications.* (pp. 37—55). Thousand Oaks, CA: Sage.

Christenson, S. L. (1992). [Review of Study Habits Evaluation and Instruction Kit]. In
J. J. Kramer & J. C. Conoley (Eds.). *The Eleventh Mental Measurements
Yearbook.* (pp. 883–884). Lincoln, Nebraska: Buros Institute.

Cizek, G. J. (2001). More unintended consequences of high-stakes testing. *Educational
Measurement: Issues and Practice, 20*(4), 19–27.

Cizek, G. J., Fitzgerald, S., Shawn, M., & Rachor, R. E. (1995). Teachers' assessment
practices: Preparation, isolation and the kitchen sink. *Educational Assessment,
3*(2), 159-179.

Clandinin, D. J., & Connelly, F. M. (1992). Teacher as curriculum maker. In P. W.
Jackson (Ed.)., *Handbook of Research on Curriculum,* (pp. 363—401). New
York: Macmillan.

Clark, C., & Peterson, P. (1986). Teachers' thought processes. In M. Wittrock (Ed.), *Handbook of research on teaching.* (3rd ed., pp. 255-296). New York: Macmillan.

Cooper, P., & Davies, C. (1993). The impact of national curriculum assessment arrangements on English teachers' thinking and classroom practice in English secondary schools. *Teaching & Teacher Education, 9*(5/6), 559-570.

Cowie, B. M. (2000). *Formative assessment in science classrooms.* Unpublished doctoral dissertation, University of Waikato, Hamilton, NZ.

Croft, A. C., & Reid, N. A. (1991). *How often and for what purposes are NZCER tests used in primary and secondary schools?* (Research Report). Wellington, NZ: NZCER.

Croft, A. C., Strafford, E., & Mapa, L. (2000). *Stocktake/evaluation of existing diagnostic tools in literacy and numeracy in English.* (Research Report to Ministry of Education). Wellington, NZ: NZCER.

Cronbach, L. J. (1970). *Essentials of psychological testing.* (3rd ed.). New York: Harper & Row.

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research, 58*(n), 438--481.

Crooks, T. J. (1990). Overview of national monitoring in education. In C. Croft (Ed.), *Proceedings of the Conference on National Monitoring in Education* (pp. 5--21). Wellington, NZ: NZCER.

Crooks, T. J. (2002). Educational assessment in New Zealand schools. *Assessment in Education: Principles Policy & Practice, 9*(2), 237—253.

Dahlin, B., Watkins, D. A., & Ekholm, M. (2001). The role of assessment in student learning: The views of Hong Kong and Swedish lecturers. In D. A. Watkins & J. B. Biggs (Eds.), *Teaching the Chinese learner: Psychological and pedagogical*

perspectives. (pp. 47-74). Hong Kong: University of Hong Kong, Comparative Education Research Centre.

Delandshere, G., & Jones, J. H. (1999). Elementary teachers' beliefs about assessment in mathematics: A case of assessment paralysis. *Journal of Curriculum and Supervision, 14*(3), 216–240.

Devine, T. G. (1991). Studying: skills, strategies, and systems. In J. Flood, J. M. Jensen, D. Lapp, & J. R. Squire (Eds.) *Handbook of Research on Teaching the English Language Arts*. (pp. 743–753). New York: Macmillan.

Dixon, H. (1999). *The effect of policy on practice: An analysis of teachers' perceptions of school based assessment practice*. Unpublished master's thesis, Massey University, Albany, NZ.

Dubois, N. F., & Kiewra, K. A. (1996). Introduction to special issue on studying. *Educational Psychology Review, 8*(4), 329–333.

Dwyer, C. A., & Villegas, A. M. (1993). *Guiding conceptions and assessment principles for The Praxis Series: Professional assessments for beginning teachers.* (Research Report RR-93-17). Princeton, NJ: Educational Testing Service.

Eisner, E. W., & Vallance, E. (1974). *Conflicting conceptions of curriculum.* Berkeley, CA: McCutchan.

Ekeblad, E. & Bond, C. (1994). The nature of a conception: Questions of context. In R. Ballantyne & C. Bruce (Eds.). *Phenomenography: Philosophy and practice*. (pp. 343-353). Brisbane, Aus: Queensland University of Technology, Centre for Applied Environmental and Social Education Research.

Elley, W. B., & Livingston, I. D. (1972). *External examinations and internal assessments: Alternative plans for reform.* Wellington, NZ: NZCER.

English, F. W. (2002). On the intractability of the achievement gap in urban schools and the discursive practice of continuing racial discrimination. *Education & Urban Society, 34*(3), 298—311.

Entwistle, N. (1997). Contrasting perspectives on learning. In F. Marton, D. Hounsell, & N. Entwistle (Eds.). *The experience of learning: Implications for teaching and studying in higher education.* (2nd ed.). (pp. 3–22). Edinburgh: Scottish Academic Press.

Entwistle, N. (n.d.). *Scoring key for the approaches and study skills inventory for students (ASSIST)*. Edinburgh: University of Edinburgh.

Entwistle, N., & Marton, F. (1984). Changing conceptions of learning and research. In F. Marton, D. Hounsell, & N. Entwistle (Eds.). *The experience of learning.* (pp. 211–231) Edinburgh: Scottish Academic Press.

Firestone, W. A., Mayrowetz, D., & Fairman, J. (1998). Performance-based assessment and instructional change: The effects of testing in Maine and Maryland. *Educational Evaluation and Policy Analysis, 20*(2), 95–113.

Fiske, E. B., & Ladd, H. F. (2000). *When schools compete: A cautionary tale.* Washington, DC: Brookings Institution Press.

Gable, R. K., & Wolf, M. B. (1993). *Instrument development in the affective domain: Measuring attitudes and values in corporate and school settings*. (2nd ed.). Boston, MA: Kluwer Academic Publishers.

Gall, M. D., Gall, J. P., Jacobsen, D. R., & Bullock, T. L. (1990). *Tools for learning: A guide to teaching study skills.* Alexandria, VA: ASCD.

Gao, L., & Watkins, D. A. (2002). Conceptions of teaching held by school science teachers in P.R. China: Identification and cross-cultural comparisons. *International Journal of Science Education, 24*(1), 61-79.

Garcia, E. (1987, April 20-24). *An ethnographic study of teachers' implicit theories on evaluation.* Paper presented at the Annual Meeting of the American Educational Research Association., Washington, DC.

Gibson, S., & Dembo, M. H. (1984). Teacher efficacy: A construct validation. *Journal of Educational Psychology, 76*(4), 569-582.

Gipps, C. V. (1994). *Beyond testing: Towards a theory of educational assessment.* London: Falmer Press.

Gipps, C. (2002). *Discussant presentation—Round table on school-based assessment.* Paper presented to the International Association for Educational Assessment (IAEA) Annual Conference, September 1 – September 6, 2002, Hong Kong.

Gipps, C., Brown, M., McCallum, B., & McAlister, S. (1995). *Intuition or evidence? Teachers and national assessment of seven-year-olds*. Buckingham, UK: Open University Press.

Glasswell, K., Parr, J., & Aikman, M. (2001). *Development of the asTTle writing assessment scoring rubrics: Report and recommendations*. Tech. Rep. 6. Auckland, NZ: University of Auckland, Project asTTle. Retrieved September 20, 2002, from http://www.tki.org.nz/r/asttle/pdf/tech_rep_06_glasswell.pdf

Goodson, I. (1995). *The making of curriculum: Collected essays.* (2nd ed.). Washington, DC: Falmer Press.

Gow, L., & Kember, D. (1993). Conceptions of teaching and their relationship to student learning. *British Journal of Educational Psychology, 63*, 20–33.

Graham, J. W., Taylor, B., J., & Cumsille, P. E. (2001). Planned missing-data designs in analysis of change. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change.* (pp. 335–353). Washington, DC: APA.

Green, K. E., & Stager, S. F. (1986). Measuring attitudes of teachers toward testing.

 Measurement and Evaluation in Counseling and Development, 19, 141-150.

Gronlund, N. E., & Linn, R. L. (1990). *Measurement and evaluation in teaching.* (6th

 ed.). New York: Macmillan.

Guskey, T. R., & Passaro, P. D. (1994). Teacher efficacy: A study of construct

 dimensions. *American Educational Research Journal, 31*, 627-643.

Guthrie, J. T. (2002). Preparing students for high-stakes test taking in reading. In A. E.

 Farstrup & S. J. Samuels (Eds.). *What research has to say about reading

 instruction* (3rd ed.), (pp. 370—391). Newark, DE: International Reading

 Association.

Hair, J. F., Jr., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate

 data analysis.* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Hall, B. W., Carroll, D., & Comer, C. B. (1988). Test use among classroom teachers

 and its relationship to teaching level and teaching practices. *Applied Measurement

 in Education, 1*(2), 145–156.

Hall, C. (2000). National Certificate of Educational Achievement: Issues of reliability,

 validity and manageability. *New Zealand Annual Review of Education, 9*, 173–

 196.

Hambleton, R. K., & Slater, S. C. (1997). *Are NAEP executive summary reports

 understandable to policy makers and educators?* CSE Tech. Rep. 430. Los

 Angeles: National Center for Research on Evaluation, Standards, and Student

 Testing, Graduate School of Education & Information Studies, University of

 California, Los Angeles.

Harker, R. K. (1982). Bourdieu and education. *New Zealand Cultural Studies Working

 Group Newsletter*. 4, 37-49.

Harootunian, B., & Yarger, G. P. (1981). *Teachers' conceptions of their own success: Current issues.* Washington, DC: ERIC Clearinghouse on Teacher Education.

Hativa, N. (2000). Teacher thinking, beliefs, and knowledge in higher education: An introduction. *Instructional Science, 28*, 331-334.

Hattie, J., Biggs, J., & Purdie, J. (1996). Effects of learning skills interventions on student learning: A meta-analysis. *Review of Educational Research, 66*(2), 99–136.

Hattie, J. (2002). *Schools Like Mine: Cluster Analysis of New Zealand Schools.* Tech. Rep. 14. Auckland, NZ: University of Auckland, Project asTTle.

Hawe, E. M. (2000). *Assessment in a pre-service teacher education programme: The rhetoric and the practice*. Unpublished doctoral dissertation, University of Auckland, Auckland, NZ.

Heaton, J. B. (1975). *Writing English language tests.* London: Longman.

Henson, R. K., Kogan, L. R., & Vacha-Haase, T. (2001). A reliability generalization study of the Teacher Efficacy Scale and related instruments. *Educational & Psychological Measurement, 61*(3), 404-420.

Herman, J. L., & Golan, S. (1993). The effects of standardized testing on teaching and schools. *Educational Measurement: Issues and Practice, 12*(4), 20-25, 41-42.

Hershberg, T. (2002). Comment. In D. Ravitch (Ed.), *Brookings Papers on Education Policy: 2002.* (pp. 324-333). Washington, DC: Brookings Institution Press.

Hian, C. H. (1994). *Operationalization and prediction of conceptions of teaching in adult education.* Unpublished doctoral dissertation, University of British Columbia, Vancouver, BC.

Hill, J. (1998). *Rethinking assessment: The challenging issues for schools and teachers.* Unpublished master's thesis, Massey University, Albany, New Zealand.

Hill, M. F. (2000a). *Remapping the assessment landscape: primary teachers reconstructing assessment in self-managing schools.* Unpublished doctoral dissertation, University of Waikato, Hamilton, NZ.

Hill, M. F. (2000b). Dot, slash, cross: How assessment can drive teachers to ticking instead of teaching. *set: Research information for teachers,* (1), 21-25.

Ho, A., Watkins, D., & Kelly, M. (2001). The conceptual change approach to improving teaching and learning: An evaluation of a Hong Kong staff development programme. *Higher Education, 42*, 143-169.

Howe, M. J. A. (1998). *Principles of abilities and human learning.* Hove, UK: Psychology Press.

Hoyle, R. H. (1995). The structural equation modeling approach: Basic concepts and fundamental issues. In R. H. Hoyle, (Ed.). *Stuructural equation modeling: Concepts, issues, and applications.* (pp. 1–15). Thousand Oaks, CA: Sage.

Impara, J. C., Divine, K. P., Bruce, F. A., Liverman, M. R., & Gay, A. (1991). Does interpretive test score information help teachers? *Educational Measurement: Issues and Practice, 10*(4), 16–18.

Jackson, P. F., Reid, N. A., & Croft, A. C. (1979). *Study habits evaluation and instruction kit.* Wellington: New Zealand Council for Educational Research.

Jensen, J. W., Kauchak, D., & Rowley, M. L. (2001, March 1-4). *Teacher candidates' beliefs: Implications for practice.* Paper presented at the 53rd Annual Meeting of the American Association of Colleges for Teacher Education, Dallas, TX.

Kahn, E. A. (2000). A case study of assessment in a grade 10 English course. *The Journal of Educational Research, 93*(5), 276-???.

Keightley, J. (2002). *School-based assessment in a high-stakes environment.* Paper
     presented to the International Association for Educational Assessment (IAEA)
     Annual Conference, September 1 – September 6, 2002, Hong Kong.

Kember, D. (1997). A reconceptualisation of the research into university academics'
     conceptions of teaching. *Learning and Instruction, 7*(3), 255-275.

Kember, D., & Kwan, K.-P. (2000). Lecturers' approaches to teaching and their
     relationship to conceptions of good teaching. *Instructional Science, 28*(5-6), 469-
     -490.

Kiewra, K. A. (1992). [Review of Study Habits Evaluation and Instruction Kit]. In J. J.
     Kramer & J. C. Conoley (Eds.). *The Eleventh Mental Measurements Yearbook*
     (pp. 884–885). Lincoln, Nebraska: Buros Institute.

Kohn, A. (1999). *The schools our children deserve: Moving beyond traditional
     classrooms and "tougher standards".* New York: Houghton Mifflin.

Koretz, D. (2002). Comment. In D. Ravitch (Ed.), *Brookings Papers on Education
     Policy: 2002.* (pp. 315-323).

Limbrick, L., Keenan, J., Girven, A. (2000). *Mapping the English curriculum.* Tech.
     Rep. 6. Auckland: University of Auckland, Project Assessment Tools for
     Teaching and Learning.

Linn, R. L. (2000). Assessments and accountability. *Educational Researcher, 29,*(2),
     4–16.

MacCallum, R. C. (1995). Model specification: Procedures, strategies, and related
     issues. In R. H. Hoyle, (Ed.). *Sturctural equation modeling: Concepts, issues,
     and applications.* (pp. 16--36). Thousand Oaks, CA: Sage.

MacKechnie, R., & MacKechnie, C. (1999). *How helpful are introductory study skills
     programmes? Thoughts on the experience of one programme in Hamilton, New*

Zealand. Paper presented at the HERDSA Annual International Conference, Melbourne, Australia, 12–15 July, 1999.

Marton, F. & Saljo, R. (1976). On qualitative differences in learning. I – outcome and process. *British Journal of Educational Psychology, 42*, 4-11.

Marton, F. (1981). Phenomenography – Describing conceptions of the world around us. *Instructional Science, 10,* 177–200.

Marton, F. (1983). Beyond individual differences. *Educational Psychology, 3*(3 & 4), 289–303.

Maruyama, G. M. (1998). *Basics of structural equation modeling.* Thousand Oaks, CA: Sage.

Masui, C., & De Corte, E. (1999). Enhancing learning and problem solving skills: Orienting and self-judging, two powerful and trainable learning tools. *Learning and Instruction, 9* (6), 517–542.

McMillan, J. H. (2001a). *Classroom assessment: Principles and practice for effective instruction.* (2nd. ed.). Boston, MA: Allyn & Bacon.

McMillan, J. H. (2001b). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice, 20*(1), 20-32.

McMillan, J.H., Myran, S., & Workman, D. (2002). Elementary teachers' classroom assessment and grading practices. *The Journal of Educational Research, 95*(4), 203-213.

Meagher-Lundberg, P., & Brown, G. T. L. (2001). *Item signature study 1.* Tech. Rep. 12. Auckland: University of Auckland, Project Assessment Tools for Teaching and Learning. Retrieved September 20, 2002, from http://www.tki.org.nz/r/asttle/pdf/tech_rep_12_meagher.pdf

Mehrens, W. A., & Lehmann, I. J. (1984). *Measurement and evaluation in education and psychology*. (3rd Edn.). New York: Holt, Rinehart and Winston.

Mertler, C. A. (1999, October 13-16). *Teachers' (mis)conceptions of classroom test validity and reliability*. Paper presented at the Annual Meeting of the Mid-Western Educational Research Association, Chicago.

Ministry of Education. (1993). *The New Zealand curriculum framework: Te anga marautanga o Aotearoa*. Wellington, NZ: Learning Media.

Neill, M. (1997). *Testing our children: A report card on state assessment systems.* Cambridge, MA: FairTest.

Noble, A. J., & Smith, M. L. (1994). *Old and new beliefs about measurement-driven reform: "The more things change, the more they stay the same".* CSE Tech. Rep. 373. Los Angeles: University of California, Los Angeles, CRESST.

Ojerinde, D. (2002). Continuous assessment practice in Africa. Paper presented to the International Association for Educational Assessment (IAEA) Annual Conference, September 1 – September 6, 2002, Hong Kong.

Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research, 62*, 307–332.

Pajares, M. F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research, 66*(4), 543-578.

Philipp, R. A., Flores, A., Sowder, J. T., & Schappelle, B. P. (1994). Conceptions and practices of extraordinary mathematics teachers. *Journal of Mathematical Behavior, 13*, 155-180.

Philippou, G., & Christou, C. (1997). Cypriot and Greek primary teachers' conceptions about mathematical assessment. *Educational Research and Evaluation, 3*(2), 140-159.

Pinar, W. F., Reynolds, W. M., Slattery, P., & Taubman, P. M. (1995). *Understanding curriculum*. New York: Peter Lang.

Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1991). *A manual for the use of the motivated strategies for learning questionnaire*. Ann Arbor, Mich: National Center for Research to Improve Postsecondary Teaching and Learning, University of Michigan.

Plake, B. S., & Impara, J. C. (1997). Teacher assessment literacy: What do teachers know about assessment? In G. D. Phye (Ed.), *Handbook of classroom assessment*. (pp. 55-68). New York: Academic Press.

Plake, B. S., Impara, J. C., & Fager, J. J. (1993). Assessment competencies of teachers: A national survey. *Educational Measurement: Issues and Practice, 12*(4), 10–12, 39.

Popham, W. J. (2000). *Modern educational measurement: Practical guidelines for educational leaders*. (3rd ed.). Boston: Allyn & Bacon.

Porter, A., & Chester, M. (2002). Building a high-quality assessment and accountability program: The Philadelphia example. In D. Ravitch (Ed.), *Brookings Papers on Education Policy: 2002*. (pp. 285-315.). Washington, DC: Brookings Institution Press.

Pratt, D. D. (1992a). Conceptions of teaching. *Adult Education Quarterly, 42*(4), 203-220.

Pratt, D. D. (1992b). Chinese conceptions of learning and teaching: A westerner's attempt at understanding. *International Journal of Lifelong Education, 11*(4), 301-319.

Pratt, D. D. (1997). Reconceptualizing the evaluation of teaching in higher education. *Higher Education, 34*, 23-44.

Pratt, D. D., & Associates. (1998). *Five perspectives on teaching in adult and higher education.* Malabar, FL: Krieger, Publishers.

Pratt, D. D., & Collins, J. B. (1998). *Teaching Perspectives Inventory.* [On-line]. Retrieved March 15, 2001, from http://www.edst.educ.ubc.ca/DPtpi.html

Pratt, D. D., & Collins, J. B. (2001, June). *The Teaching Perspectives Inventory (TPI).* Paper presented at the Adult Education Research Conference, Vancouver, BC.

Pressley, M., Yokio, L., van Meter, P., Van Etten, S., & Freebern, G. (1997). Some of the reasons why preparing for exams is so hard: What can be done to make it easier? *Educational Psychology Review, 9*(1), 1–38.

Prosser, M., & Trigwell, K. (1999). *Understanding learning and teaching: The experience in higher education.* Ballmoor, UK: SRHE & Open University Press.

Purdie, N., & Hattie, J. (1999). The relationship between study skills and learning outcomes: A meta-analysis. *Australian Journal of Education 43,* (1), 72–86.

Quilter, S. M. (1998). *Inservice teachers' assessment literacy and attitudes toward assessment.* Unpublished doctoral dissertation, University of South Carolina, Columbia, SC.

Quilter, S. M., & Chester, C. (1998, 14-17 October). *Inservice teachers' perceptions of educational assessment.* Paper presented at the Annual Meeting of the Mid-Western Educational Research Association, Chicago.

Resnick, L. B., & Resnick, D. P. (1989). *Assessing the thinking curriculum: New tools for educational reform.* Washington, DC: National Commission on Testing and Public Policy.

Robinson, V., Phillips, G., & Timperley, H. (in press). Using achievement data for school-based curriculum review.

Rohwer, W. D., Jr. (1984). An invitation to an educational psychology of studying. *Educational Psychologist, 19*(1), 1–14.

Rosenshine, B., Meister, C., & Chapman, S. (1996). Teaching students to generate questions: A review of intervention studies. *Review of Educational Research, 66* (2), 181–221.

Rotter, J. B. (1982). Social learning theory. In N. T. Feather (Ed.). *Expectations and actions: Expectancy-value models in psychology* (pp. 241–260). Hillsdale, NJ: Erlbaum.

Sadler, R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*, 119-144.

Saltzgaver, D. (1983). One teacher's dominant conceptions of student assessment. *Curriculum Perspectives, 3*(2), 15-21.

Samuelowicz, K. (1994). Teaching conceptions and teaching practice: A case of assessment. In R. Ballantyne & C. Bruce (Eds.). *Phenomenography: Philosophy and practice*. (pp. 343-353). Brisbane, Aus: Queensland University of Technology, Centre for Applied Environmental and Social Education Research.

Samuelowicz, K., & Bain, J. D. (1992). Conceptions of teaching held by academic teachers. *Higher Education, 24*, 93-111.

Schacter, D. L. (1999). The seven sins of memory: Insights from psychology and cognitive neuroscience. *American Psychologist, 54,* (3), 182–203.

Schmeck, R. R. (2000). Thoughtful learners: Students who engage in deep and elaborative information processing. In R. J. Riding & S. G. Rayner (Eds.). *International perspectives on individual differences, vol. 1.* Stamford, CT: Ablex Publishing..

Schmeck, R. R., Ribich, F, & Ramanaiah, N. (1977). Development of a self-report inventory for assessing individual differences in learning processes. *Applied Psychological Measurement, 1*(3), 413–431.

Schoenfeld, A. H. (2002). Making mathematics work for all children: Issues of standards, testing, and equity. *Educational Researcher, 31*(1), 13–25.

Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist, 54*(2), 93–105.

Scriven, M. (1991). Beyond formative and summative evaluation. In M. W. McLaughlin & D. C. Phillips (Eds.), *Evaluation & education: At quarter century* (Vol. Part II, pp. 19–64). Chicago: NSSE.

Senk, S. L., Beckmann, C. E., & Thompson, D. R. (1997). Assessment and grading in high school mathematics classrooms. *Journal for Research in Mathematics Education, 28*(2), 187–215.

Smith, M. L. (1991). Meanings of test preparation. *American Educational Research Journal, 28,* 521–542.

Smith, M. L., & Fey, P. (2000). Validity and accountability in high-stakes testing. *Journal of Teacher Education, 51*(5), 334–344.

Smith, M. L., & Rottenberg, C. (1991). Unintended consequences of external testing in elementary schools. *Educational Measurement: Issues and Practice, 10*, 7–11.

Smith, M. L., Heinecke, W., & Noble, A. J. (1999). Assessment policy and political spectacle. *Teachers College Record, 101*(2), 157–191.

SPSS. (2000). SPSS for Windows. (Version 10.1.0) [Computer software]. Chicago, Il: SPSS.

Stamp, D. (1987). *Evaluation of the formation and stability of student teacher attitudes to measurement and evaluation practices.* Unpublished doctoral dissertation, Macquarie University, Sydney, Australia.

Stiggins, R. J. (2001). The unfulfilled promise of classroom assessment. *Educational Measurement: Issues and Practice, 20*(3), 5-15.

Sturrock, F. (1999). *Teacher census: Preliminary report.* Unpublished Report. Wellington, NZ: Ministry of Education, Demographic and Statistical Analysis Unit.

Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology.* San Francisco: Jossey-Bass Publishers.

Thompson, A. G. (1992). Teachers' beliefs and conceptions: A synthesis of the research. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning.* (pp. 127–146). New York: Macmillan.

Thorndike, R. M. (1997). *Measurement and evaluation in psychology and education.* 6th Edition. Upper Saddle River, NJ: Prentice-Hall.

Timperley, H, & Robinson, V. (2002). *Partnership: Focusing the relationship on the task of school improvement.* Wellington, NZ: NZCER.

Tittle, C. K. (1994). Toward an educational psychology of assessment for teaching and learning: Theories, contexts, and validation arguments. *Educational Psychologist, 29*(3), 149-162.

Torrance, H., & Pryor, J. (1998). *Investigating formative assessment: Teaching, learning and assessment in the classroom*. Buckingham, UK: Open University Press.

Trigwell, K., & Prosser, M. (1997). Towards an understanding of individual acts of teaching and learning. *Higher Education Research & Development, 16*(2), 241–252.

Tschannen-Moran, M., Woolfolk Hoy, A., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research, 68*(2), 202–248.

Vermetten, Y. J., Vermunt, J. D., & Lodewijks, H. G. (1999). A longitudinal perspective on learning strategies in higher education: Different viewpoints towards development. *British Journal of Educational Psychology, 69,* 221–242.

Waltmann, K. K., & Frisbie, D. A. (1994). Parents' understanding of their children's report card grades. *Applied Measurement in Education, 7*, 223-240.

Warren, E., & Nisbet, S. (1999, 4-7 July). *The relationship between the purported use of assessment techniques and beliefs about the uses of assessment.* Paper presented at the 22nd Annual Conference of the Mathematics Education and Research Group of Australasia, Adelaide, SA.

Webb, N. L. (1992). Assessment of students' knowledge of mathematics: Steps toward a theory. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning.* (pp. 661–683). New York: Macmillan.

Weinstein, C. E., & Palmer, D. R. (1990). *Learning and study strategies inventory-high school version.* Clearwater, FLA: H & H Publishing Co.

Wiliam, D., & Black, P. (1996). Meanings and consequences: A basis for distinguishing formative and summative functions of assessment? *British Educational Research Journal, 22*(5), 537–548.

Wylie, C. (1997). *Self-managing schools seven years on: What have we learnt?* Wellington: NZCER.

Zimmerman, B. J. (2001). Theories of self-regulated learning and academic

achievement: An overview and analysis. In B. J. Zimmerman & D. H. Schunk

(Eds.), *Self-regulated learning and academic achievement: Theoretical*

perspectives. (2nd ed., pp. 1–36). Mahwah, NJ: LEA.