

Chapter 2

Evaluation

Introduction

Evaluation has had a chequered background. It has often been responded to with suspicion by teachers, generated misinformation from both teachers and management, and been subject to misdirection by administrators or bureaucrats. This chapter reviews the history of evaluation, examines evaluation methodologies, and considers the factors and influences that bear on this topic. This analysis ultimately yields a list of essential elements for evaluation that both illuminate past failures, and provides a way forward for more successful implementation of this important process for the enhancement of teacher quality and consequently student learning.

Whilst evaluation may not be a common term in Tasmanian teacher circles, it does have a historical context that continues, either through knowledge or suspicion, to impact on teacher attitudes and perceptions. The experiences in England and the United States may differ from Australia, but these are related to reveal a common desire to identify measures of teacher quality by various means. The historical sections of this chapter illustrate that, where initiatives have failed to address some of the fundamental concerns of teachers or been misdirected, such as in England, the cause of teacher evaluation has been affected adversely so that there is little ground for optimism. Conversely, teacher owned initiatives in the United States such as the National Board of Professional Teaching Standards, reveal new and positive approaches for teachers to evaluate their practice. This chapter also cautions that pressures for productivity and dictated social outcomes continue to threaten judgement of schools and teachers based on standardised test scores which lack validity, and which fail to appreciate factors coming to bear on student results as well as the broader requirements of a holistic education.

Because teachers work in schools that deliver mass education, organisation theory is introduced to determine if there are inherent factors that limit and hinder the application of evaluation to teachers. The work of Bidwell (1965) and Pusey (1976) are drawn on to illustrate unique aspects of education as compared to other enterprises. These theorists demonstrate that schools deliver an uncertain product under unusual conditions of isolation, and indeterminate bureaucratic control of the teaching act itself. Elmore's (2000) more recent concept of "loose coupling" explains why schools themselves may hinder the development of instruments of surveillance of teacher practice. Together, these researchers help explain why evaluation might

not emerge in a more natural fashion in the educational enterprise, particularly when hierarchical or bureaucratic agendas exceed the simple concept of student learning.

Against this background, the issues flowing from established evaluation methodologies are drawn out from the literature, after which discussion shifts to the purposes behind, or motives for, evaluation. Organisationally dominant appraisal models are questioned because of their inherent hostility that causes teachers to become defensive and quarantine their practice. Ritualistic evaluation processes, or those that are directed to identify culprits for systemic failures, obviously fail to survive because of teacher cynicism or inappropriate targeting. In particular, the question of competence, or conversely incompetence, is presented and the question raised as to whether such a judgement ever can be made objectively. This leads to a consideration of the theoretical purposes of evaluation, and a clear statement that inconsistent or inaccurate objectives yield imperfect, and also destructive, results. Quite simply, if those who implement a system are unclear about why they are doing so, the outcomes of confusion or suspicion seem like natural consequences, whereas the intended purpose of enhanced teacher quality becomes elusive.

Evaluation approaches, such as standards, duties, and peer and self-appraisal, are described, yielding the eventual judgement that the use of either is appropriate to the right purpose, but one must accept the limitations of the model selected. The concept of the teacher as reflective practitioner as defined by Schon (1983) is presented to involve individual practitioners in a debate that often has occurred without direct teacher involvement. The factors covered in this section restate the importance of identifying an appropriate purpose to which a particular model may be attached to achieve a pre-considered outcome. To mistake this clear relationship is to invite difficulties. By this stage, the chapter will have established the fundamental issues that have made evaluation complex and rather difficult to achieve in reality, although it need not have been so.

The focus then shifts to additional factors, specifically the role of change, and the issue of professionalism. Because evaluation contains within it the concept of change for the better, and concurrently the educational landscape is varying in response to research and societal needs, the management of change and consequent disequilibrium become significant to this topic. Teachers who feel insecure need support and encouragement, and it is questioned whether evaluation processes introduced to date have taken this into account sufficiently. The matter of professionalism focuses the issues of this section away from observation of external

factors to consider whether teachers themselves have a role to play in evaluation, and whether they have to date engaged with the topic in a professional manner.

The discussion of the various issues and factors impacting on the development of appropriate evaluation procedures ultimately generates a synthesised list of design principles for evaluation. This list both provides a starting point for both the quantitative and qualitative research in this paper, and establishes some common principles that would optimise the chance of evaluation being implemented successfully to the benefit of student learning and teacher professional identity.

Historical Context

The history of evaluation from old style school inspections to the present has been marked primarily by divergent purposes and intermittent attempts at performance appraisal. Partisan politics and sensationalism about declining educational standards and opportunities have had greater coverage than the quality of learning (Wells 1989, p.25; Senate Employment & Training References Committee 1998, p.42; Ramsey 2000, p.126). The size and complexity of educational processes in modern society, and the demands of extensive economic changes since the Second World War have ensured the continuation of calls for productivity, accountability, and improved educational standards (Istance & Lowe 1991, pp.24-25; Elkins & Elliott 2004, pp.16-18; Hodkinson & Hodkinson 2005, p.111; Parr 2004, p.11). Such pressures probably will continue whilst post-Keynesian economic rationalist governments emphasise education as a factor in meeting national economic objectives whilst concurrently attempting to reduce their level of participation in social programs such as education (Scott et al. 2001; Angus 1991, p.148; Smyth et al. 1997, p.13; Tickle 2000, p.12).

United States

Despite some earlier work on standardised achievement testing for the measurement of outcomes, the primary impetus for implementation of teacher evaluation in the United States was a combination of post-war social and technological change (Worthen & Sanders 1973, p.2). By the late 1950's, the "baby boomers" born in the post-war reconstruction years headed into schools in an environment where 'There was a positive conviction that education was a positive good - for an individual the route to social mobility; for society, the motor of prosperity' (Istance & Lowe 1991, p.22). Expanded education budgets and a perceived strategic and technological gap with the Soviet Union prompted by the launch of *Sputnik* soon translated into calls for accountability (Istance & Lowe 1991, p.31).

Debate in the US Congress on the Elementary and Secondary Education Act of 1965 raised concerns as to the verifiability of the thousands of federal grants being put to appropriate uses, and Senator Robert F Kennedy among others argued for an accountability system where educators were required to evaluate their own efforts (Worthen & Sanders 1973, pp.4-5). Worthen and Sanders note that, despite this demand, 'Few scholars had turned their attention to the development of generalizable evaluation plans which could be adopted or adapted by local evaluators. Theoretical work in evaluation was almost non-existent' (Worthen & Sanders 1973, p.6). The contributors to Worthen and Sanders' work attempted to fill this gap through the application of system theory that was predominant in the commercial and political area at the time. They illustrated their discussion of programs with flow charts and statements like 'Evaluation is the watchdog of program management' (Provus, 1973, p.186). The way was opened in the United States for legislators and the public to question not only financial usage but also student outcomes as the measure of educational effectiveness (Poster et al. 1991, p.22).

Researchers saw the changes of the 1970's as significant in the development of evaluation practices in the USA. The promise of the 1960's, and the predictions of experts, failed to materialise (Schon 1983, p.11; Lokan and McKenzie 1989, p.2; Freeland 1991, p.65). Unemployment and difficulties in entering the labour market 'all caused some people to blame schools for failing to prepare [pupils] adequately for working life' (Istance and Lowe, 1991, p.23. See also Ball & Goodson 1985, p.5). Poster, Poster and Bennington suggest that an almost imperceptible transition occurred in this period away from the elements of complex evaluation of innovation to a focus on outcomes (Poster et al. 1991, p.22). Studies of the time advanced the view that, if student outcomes could be influenced by good teaching, then it should be feasible to encourage this by rewarding effective teachers and penalising others (Lokan and MacKenzie 1989, p.2). By the end of the 1970's compulsory evaluation had been introduced into a number of American states (Poster et al. 1991, pp.22-33). Public perception had placed improving teacher quality ahead of improvements in curriculum and school management as the primary method for educational reform (Darling-Hammond et al. 1986, p.204).

In response to declines in results in standardised test scores, the sensationally titled *A Nation at Risk* (1983) confirmed the trend in the debate away from content and class organisation towards student outcomes, and thus teacher competence (Darling-Hammond 1988, p.57; Poster et al. 1991, p.22). The formation of the National Board of Professional Teaching Standards

(NBPTS) in 1987 represented the first step to counter the pressures from using test scores as the measure of teacher competence. However, response to standardised tests appears to remain the primary determinant of teacher effectiveness in the United States (Darling-Hammond 1997, p.56; Schalock et al. 1993, pp.121-122; Archer 2002). Meanwhile, those evaluation procedures that had been put into place focused less on improving pedagogy than on accountability and promotion selection (Loup et al. 1996, p.215).

England

England differs from the United States in that it retains a central educational inspectorate, now termed the Office for Standards in Education, generally referred to as Ofsted. Following calls for higher standards and greater accountability in education in 1976, local education authorities (LEA) in England encouraged the implementation of teacher appraisal schemes involving self-appraisal, interviews with peers, and limited classroom observation (Forde 1989, p.34). By 1985, one in four schools in southern England was operating a staff appraisal scheme with Her Majesty's Inspectors noting that 'teacher appraisal is expected to provide a more systematic, objective and reliable evaluation, identify those teachers with promotional potential and inform managerial decisions in schools and in the local Education Authority' (Forde 1989, p.38; Poster et al. 1991, p.13). Some teachers were suspicious that appraisal was a tool for reducing the teaching force at a time of falling student numbers (Poster et al. 1991, p.11). In 1986, Secretary of State Kenneth Baker proposed a link between appraisal and salaries, an issue that became part of the teachers' pay dispute of 1987 (Forde 1989, pp.35, 39). By 1990, the attempt to regulate compulsory appraisal in the UK was shelved and the focus moved from managerial control to an emphasis on appraisal as a strategy for professional development (Poster et al. 1991, pp.17-18).

One of the reasons cited for dropping compulsory appraisal was the work associated with introducing the National Curriculum. However, Elliott (1998) argues that this implementation by the Thatcher government sought not only to marginalise teachers, but also to roll back some of the progressive reforms of the 1960's. The revolt in 1993 by English teachers against Key Stage 3 standardised testing for 14 year olds slowed the trend towards an American testing style and subsequent judgement of teachers (Elliott 1998, pp.18-20). Nonetheless, by September 2001, the *Education (School Teacher Appraisal) (England) Regulations* had established mandatory appraisal for all teachers employed by local education authorities (HM Government 2001). However, in 2003 Ofsted conducted a survey of the employment of Advanced Skills Teachers, established in 1998 'to help schools attract and retain excellent

classroom teachers and to increase staff motivation, raise pupils' achievement levels and broaden the skills and knowledge base of schools' (Ofsted 2003, p.4). The inspectors reported that:

- Performance management is generally not used effectively to channel and support the ASTs' work. It is rare for performance management objectives to include reference to the distinctive features of their role or, in particular, to include specific reference to the outreach activities that they undertake.
- Performance management rarely includes an analysis of the development needs of ASTs for their specific role. Overall, opportunities for their professional development are few (Ofsted 2003, p.6).

In England, attempts to reward teachers at the lower levels through the establishment in 2000 of a performance threshold grading, for those who had completed their induction phase and were considered having achieved high standards, suffered similar difficulties. The implementation was flawed, with the performance mechanism trailing well behind the admission process. The result was that, of the 80% of the eligible cohort who applied, 97% were admitted to the threshold payment level largely unaware that an appraisal element was part of the process (Menter et al. 2002, p.4; Elkins & Elliott 2004, p.25). The threshold system had its origin in the private sector, and was even tendered out for administration by private companies (Menter et al. 2002, p.8). The management system was laborious and bureaucratic, and it engendered at least initially a compliance mentality in response to what effectively appeared to be a complex system of awarding a pay rise. Subsequent research has identified that the process had a negative impact on the well being of many participants and generally damaged teacher morale (Kleinhenz & Ingvarson 2004, p.41; Haynes et al. 2003, pp.32, 36). Researchers of unsuccessful candidates found that

Many felt an overwhelming sense of injustice. This was founded initially on the limitations of the application form to allow them to describe, in full, their knowledge, skills and contributions. Second, many felt that judgements were made on their classroom practice when this had been only rarely, or in some cases, never observed (Haynes et al. 2003, p.32).

Nonetheless, the view of some head teachers was that the differentiation between teachers would flow in the advancement after the threshold (Menter et al. 2002, p.8). Unfortunately, the

subsequent review of the Advanced Skills Teacher system mentioned above suggests that this was an optimistic view (Elkins & Elliott 2004, p.26). The research judgement of the threshold process was that there was no evidence of improved classroom practice and participants were left with a feeling that advancement was based on good record and data keeping rather than quality of teaching (Haynes et al. 2003, p.43).

Clearly, there has in the past been a gap between the regulations and reality, a situation that may well continue. What the centralised regulations and functions of Ofsted reveal is a continuing determination to drive the teacher quality agenda from the top down. The October 2005 white paper *Higher Standards, Better Schools for All* continues this trend, noting that:

Better standards are dependent on better teaching. The quality of teaching in our classrooms has been transformed since 1997. Ofsted judged teaching to be good or better in 78% of secondary schools inspected in 2004/05 compared with 59% in 1996/97, and in 74% of primary schools, compared with 45% in 1996/97. According to Ofsted, we already have the best ever generation of teachers...We will make performance management more effective. The greatest rewards and promotion throughout a teacher's career will go to those who make the biggest impact on pupils' progress and who show commitment to the development of themselves and their colleagues. The best training will be delivered in schools by our best teachers and we will ensure classroom observation and feedback are improved (HM Government 2005, pp.85-86).

Based on the threshold and advanced skills teacher processes, the achievement of such targets are likely to remain in the realm of political pronouncement rather than actual improvement. Kleinhenz and Ingvarson doubt whether 'validity and fairness can be attained as long as responsibility for the evaluations remains with principals, and the evaluations are carried out, as at present, at school sites' (Kleinhenz & Ingvarson 2004, p.41).

Australia

For most of its educational history, Australia followed the British lead of using inspectors to verify performance (Kleinhenz & Ingvarson 2004, p.33). In the 1860's a British scheme to pay teachers based on the level of attendance and testing against standards in reading, writing and arithmetic was instituted in the colonies (Phillips 1985, p.48; Rapple 1994). In Tasmania at least, these regulations lasted for only three years. Not only was it hard to convince parents to

release their children from farm work to school, but also the regulations were a disincentive to teach in a state that already had staff shortages.

In most cases the inspectors acted as watchdogs, generating fear and hostility in both schools and teachers (Preston 1989, p.19). However, they were neither perfect themselves, nor were they utilised in a punitive fashion. William Lewis Neale, an Inspector of Schools in South Australia, was commissioned to report on the condition of primary education in Tasmania in 1904. Neale criticised the Tasmanian inspectors for

the fact that for 350 of the 500 State school teachers there were no records kept of their conduct, skill, and efficiency. As a consequence, promotions and transfers were made on personal knowledge. In their examination of the children, Inspectors did not know what they were examining as no standard of attainment had been agreed upon between them (Phillips 1985, p.87).

As Director of Education in Tasmania from 1905 to 1909, Neale attempted to use inspectors to enhance teaching effectiveness in a more enlightened fashion. He encouraged inspectors to move away from detailed scrutiny to acting as professional advisors to head teachers. This was continued by Neale's successor William Taylor McCoy until his departure in 1919 (Phillips 1985, p.135). The sixth Director of Education in Tasmania, George Vickery Brooks, in 1923

asked teachers to exercise their own judgement, to have confidence in themselves as professional people, to rely without reservation upon the quality of their training, to experiment with method and to be assured that, in doing do, they would have the support of the Director and Inspectors (Phillips 1985, p.172).

Throughout Australia since the ending of the inspection system, teachers have had no experience of formal scrutiny of their work unless they sought promotion or were accused of incompetence, but 'Even then, this rarely involved first hand evidence of teacher performance, or their students' work' (Kleinhenz & Ingvarson 2004, p.33. See also Kleinhenz et al. 2002, p.3). Attempts were made to rectify this situation. For example, the 1985 Quality Education (Teacher Efficiency) Review in NSW was to be a 'thorough review of the means by which efficiency of the NSW Teaching Service is monitored and assessed. Procedures for the expeditious dismissal of inefficient teachers will form part of the review' (Preston 1989, p.20). However, by April 1986 this was abandoned in the face of teacher opposition. Award restructuring under the National Wage Case decisions of 12 August 1988 sought to reward

better teachers and keep them in the classroom, and stem the flow into administrative positions in pursuit of higher salaries (McRae 1994, pp.144-145). The government sponsored Schools Council identified evaluation of teachers' work as a stage of microeconomic reform and award restructuring. It supported the establishment of the Advanced Skill Teacher (AST) scale to not only improve morale, but also raise productivity (Schools Council 1989, pp.53-54). The report *Teacher Quality* mentioned teacher appraisal but failed to provide any detail of the form or structure of such a process. The recommendations were bureaucratic and vague, such as:

the issue of teacher appraisal may be central to the question of how teachers will move to the new scale...Whatever forms of teacher appraisal are ultimately developed through the processes of negotiation between teachers and their employers in the context of award restructuring, it is hoped that the outcomes provide an important impetus to the improvement of teachers' career development prospects (Schools Council 1989, pp.55, 57).

The Schools Council's subsequent report *Australia's Teachers: An Agenda for the Next Decade* criticised appraisal practices in existence at the time as 'cursory, lacking credibility, uninformative, unhelpful and associated with a form of control devoid of responsibility for the teachers' actual performance' (Schools Council 1990, p.96). Once again the report does not go into detail of the structure of such appraisals, although specific reference is made to unions working against them. *Australia's Teachers* suggested that teachers should be required to undergo a formal appraisal every five years and more frequently if they so wished (Schools Council 1990, p.98). The process was to "sustain the process of teachers' professional growth", include the teacher in the design process, and respect the teacher's judgement about who should be involved in the appraisal, be it peers, the principal, or other experts (Schools Council 1990, pp.98-99). Appraisal for incompetence was mentioned, but heavily qualified with broad terms: 'On the basis of sound evidence, with the full knowledge of the teacher concerned and following prior discussion, a school leader should have the capacity to *require* an assessment of a teacher believed to be professionally incompetent' (Schools Council 1990, p.102).

Whilst it went further than its predecessor, *Australia's Teachers* was heavily qualified, not only in acknowledging the cost and training burden of establishing an evaluation system but also in the tone of its language: 'However, even the introduction and encouragement of more informal and unsystematic forms of appraisal would be better than no action in this area at all' (Schools Council 1990, p.97). It supported appraisal as the basis for Advanced Skills Teacher (AST)

categorisation, but provided no guidance on how this was to occur. Smyth and others consider that a critical flaw in the process was established by the differing views of teachers who favoured a more narrative portrayal of skills, and bureaucrats who required lists of predetermined traits and behaviours (Smyth et al. 1997, p.14). Difficulties in resolving such issues meant that, when the first group of AST candidates was submitted in 1991, 90% were approved with little other than application forms which failed to indicate what were advanced skills between subjects or teaching levels (Chadbourne & Ingvarson 1994, pp.273-274). *Advanced* came to be interpreted as *well above average*. Pay for better work became a marginal pay increase for different work. Chadbourne and Ingvarson criticise the AST process as a lost opportunity:

AST 1 was not intended to be a payment-by-seniority position. It was doubtful whether teachers or the public believed that being “on the top scale for a long time” tells us anything about the quality of teacher’s skills (Chadbourne & Ingvarson 1994, p.275. See also: Ingvarson 1995, pp.35-36; McRae 1994, pp.144-145; Kleinhenz & Ingvarson 2004, p.34).

Despite the AST1 verification being limited initially, teachers still felt threatened by the need to apply and in some cases face interviews to justify themselves and their pedagogy before an official panel, making the process negative rather than affirming, and consequently a terminal rather than initiating process for teacher quality (Smyth et al. 1997, p.16). The AST 2 and 3 categories which replaced the senior teacher classification were subjected to a quota and tied to specific jobs, further undermining the concept of rewarding advanced teaching performance (Preiss 1994, p.267).

Research in 2001 identified that all state education systems in Australia recommended formative appraisal of new teachers and that this was to be followed by a summative evaluation prior to transition to full employment. Furthermore there was an ‘expectation in all Australian state government schools that teachers will go through processes of regular performance appraisal which are the responsibility of the principal’ (Kleinhenz et al. 2001). However, the criteria were superficial and appear to have been more confined to the pages of departmental publications rather than actual practice, such that the state educational systems ‘would be hard put to demonstrate that the work of teachers in this period [was] being evaluated in ways that address[ed] the core of teaching practice’ (Kleinhenz et al. 2001).

The outcome of the decline in the inspection system and subsequent ineffectual efforts is that ‘There are many, possibly even a majority of teachers working now in Australian schools who have *never*, in careers spanning over thirty years, opened up their classroom practice to any kind of formal evaluation or scrutiny’ (Kleinhenz & Ingvarson 2004, p.33).

A comparison of these three countries suggests that nowhere has the question of teacher evaluation been resolved completely, despite thirty years of discussion. The establishment of the National Board for Professional Teaching Standards in the United States is certainly a step toward redressing the balance, but continuing widespread reliance on standardised test outcomes elsewhere in the United States in the eyes of many has distorted the question of teacher effectiveness away from good student learning to test scores. Britain has developed a mix of devolved appraisal procedures responding to centralised control, so that teacher suspicion has generated compliance in order to achieve pay rises, and the rhetoric of enhanced standards appears to exceed actual verification.

Meanwhile, discussion has prevailed over action in Australia to the extent that Kleinhenz, Ingvarson and Chadbourne have described it as ‘something of a *tabula rasa* with regard to teacher evaluation’ (Kleinhenz et al. 2002, p.3). Whilst an analysis in 2004 by the same researchers of the Experienced Teacher with Responsibility process in Victoria and the Level 3 classroom teacher with responsibility evaluation in Western Australia might have suggested that progress had been made, they were nonetheless motivated to judge that the evaluation processes established in the interim reflected

wide gaps between managerially designed and implemented procedures and the realities of what teachers actually know and do...they are resulting in shallow, superficial assessments that most teachers do not take seriously...crude schemes continue to be proposed, and little really changes (Kleinhenz & Ingvarson 2004, p.37).

It is notable that in all cases teachers have either failed or been slow to initiate their own evaluation for certification. An Australian College of Education discussion paper cited the journalist Adele Horin who wrote in *The Sydney Morning Herald* in 2000 that teachers in Australia appeared to remain happy in their tenured positions, progressing up the seniority and pay scales without any evaluative procedures or checks on their quality of work (ACE 2000, p7). Meanwhile, teacher unions have not engaged with the issue, seeing evaluation as a management problem until such time as members feel their jobs or conditions threatened

(Ingvarson 1989, p139; McLaughlin & Pfeifer 1988, p1; Angus 1991, p152; Cumming & Jasman 2003, p.20). If the positive tones of the new Australian curricular documents seeking to integrate learning to meet the needs of the future are to be realised, teachers need to appreciate their work, learn to evaluate it, and utilise it to reinvent their teaching for the future. It is time to make a difference. As Ramsey wrote in 2000:

The debate of the past 20 years about standards and how to improve the quality of teacher education has run its course. It is time to move forward. Most teacher educators and teachers are now at the point where they are disillusioned by seemingly endless debate and a repetitive chain of reviews which, in spite of their findings and recommendations in such critical areas as funding, standards of professional practice, accreditation of initial teacher education programs and teacher licensing, fail them (Ramsey 2000, p.31).

Organisation Theory

The School as Bureaucracy - Weber

The hierarchical and formalistic impersonality of modern schools is almost inevitably consequent on education being a mass movement in society regulated by governments (Dimmock 1995, p.276; Ball & Goodson 1985, pp.8-9). In this they conform to Weber's definition of bureaucracy, and the characteristics flow from both the demands of the activity and the society in which education occurs. Mass education becomes a flow-on from mass democracy. The reduction of economic and social differences progresses to a stage where the education of society is essential, not just for the continuation of economic development but also as a right of the citizen who is entitled to the benefits of the products of the society (Weber 1970, p.15). The strictly rational and impersonal characteristics are consequent on the need for both efficiency and impartiality for the children of a broad range of abilities and backgrounds who are thrust together in schools (Weber 1970, pp.13-14). This impartiality is possible because bureaucracy 'succeeds in eliminating from official business love, hatred, and all purely personal, irrational, and emotional elements which escape calculation. This is the specific nature of bureaucracy and it is appraised as its special virtue' (Weber 1970, p.14. See also Perrow 1970, p.58).

Schools can be said to conform to Weber's definition in that there is a clear division of hierarchical responsibility and functional specialisation, particularly where teachers are 'set for habitual and virtuoso-like mastery of single and yet methodically integrated functions' in their subjects (Weber 1970, p.17). Principals and their deputies preside over heads of schools and

faculties. From the perspective of students and parents, teaching is delivered objectively, in accordance with established procedure (what Weber calls *calculable rules*), and ‘without regard for persons’ (Weber 1970, p.15). The bureaucratic structure expects teachers over their extended career period to deliver consistent and predictable instruction to all students equally. Where changes occur, their incorporation is also expected from teachers who are themselves in a process of continual learning (Weber 1970, p.13). It is this aspect of Weber’s bureaucracy that is seen as the most basic definition of teacher competence. It is the superficial perception of all students in front of their desks with teachers before the board whilst school administrators work in their offices ensuring the administrative balance of the structure. From a Weberian perspective, order and efficiency define success as the process of education undertaken in a *rational* fashion. ‘Precision, speed, unambiguity, knowledge of the files, continuity, discretion, unity, strict subordination, reduction of friction and of material and personal costs - these are raised to the optimum’ (Weber 1970, pp.12-13).

From Weber’s perspective, the teacher effectively enters into a contract whereby status, remuneration and security are the consequences of acceptance of specific obligations. Provision of tenure guarantees ‘a strictly objective discharge of specific office duties free from all personal considerations’ (Weber 1970, p.9). Thus, just as the expectation is that teachers will deliver services to students objectively, so the structure itself is meant to ensure the ability of teachers to operate under a hierarchy freed from the destabilising influences of charismatic or traditional leadership (Weber 1970, p.5). Superiors may come and go or change positions, but the inherent rationality of the structure remains as each member has ‘a common interest in seeing that the mechanism continues its functions and that the societally exercised authority carries on’ (Weber 1970, p.17). However, this comes at a cost. Bureaucracy is for Weber one of the hardest social structures to destroy, and within it the individual is a “cog”, locked into a specific function that cannot be changed short of progressing through the various status levels to the very top.

Bidwell’s Structural Looseness

Bidwell accepts that the rational bureaucratisation of schools was somewhat inevitable given the need for schools to establish consistent methods of assessment in order to “output” students of a minimum competence level (Bidwell 1965, p.974). The extent and duration of the process of socialising children for adult roles demands to some degree the efficiencies, ordering, and rationality of bureaucracy as espoused by Weber. Bidwell explains this as the development of ‘increasingly differentiated and demanding socialization tasks’ (Bidwell 1965, p.975). In other

words, as the retention rate has climbed, society has established greater expectations of social development, both formally through assigned content, and as a result of pressure from indirect clientele in the form of parents (Bidwell 1965, p.978).

However, Bidwell modifies Weber by identifying a series of tensions as well as what he calls *structural looseness* in the form of school operations. Firstly, the demand of the system of mass education is for the systematic organisation of students by age into grades and the matching of these to expected uniform minimum outcomes regardless of variations in particular aptitudes or levels of socialization so that:

all students, by the time they leave school, must have acquired rudimentary competence for adult citizenship, that is, for enacting roles common to all adult members of the society. Students also are to be prepared for differential roles, most notably in the occupational sphere (Bidwell 1965, p.974).

However, in a situation of compulsory education, attendance at school may well be contrary to the natural inclination of the student, and the school must accept the attendance of those it may not desire (Bidwell 1965, p.973). Thus, the maintenance of control in schooling is illusory. Rather, it is an uneasy equilibrium with the coalition of teachers “faced off” against coalitions of students; order and motivation confronting a social and personal inertia and resistance to control (Bidwell 1965, p.980). A desire to be out on the beach on a sunny day is not deviance, but a natural desire for youth. Keeping students in schoolrooms behind desks is the unnatural act. Consequently, as Perrow notes ‘a great deal of organisational effort is exerted to *control* the effects of extra-organisational influences’ (Perrow 1970, p.52).

A second tension exists between the function of the teacher as an agent of the state, expected to move students through the curriculum based on their allocated age level and curriculum grading, and the teacher as educator and server of client needs, who must appreciate and factor in to the educational process the variations between the academic competence of each student and other influences that generate differences in student learning. The establishment of interpersonal bonds between the teacher and students results not just from the time spent together. These bonds are necessary to enable the teacher to gauge variations and establish an affective framework for motivation and the facilitation of learning (Bidwell 1965, p.974). Variations in intellectual and societal development, and motivational levels of each individual student mean that the bureaucracy cannot dictate this aspect of a teacher’s work. The affective bonds, upon which teacher reconciliation of the factors is based, are alien to the bureaucratic

function where discretion and personal consideration contradicts the impersonal order that should eliminate 'personal, irrational, and emotional elements' (Weber 1970, p14. See also Bidwell 1965, p.979).

This tension is overcome by teachers being accorded a level of autonomy within the unique classroom setting for the exercise of discretionary judgement towards students in their charge. Bidwell describes a condition of *structural looseness* where 'The teacher works alone within the classroom, relatively hidden from colleagues and superiors, so that he has a broad discretionary jurisdiction within the boundaries of the classroom' (Bidwell 1965, p.976). In other words, the bureaucracy has to acknowledge that there is an area that it cannot readily penetrate or directly control. The existence of this area is tolerated, provided that the elements of personal contact that occur in the classroom are not threatening to the structural order of the school or the progression of students through the system in their cohorts (Bidwell 1965, p.1013). This structural looseness has in fact increased with the expansion of administrative structures to deal with the logistic demands of mass schooling and greater focus in teaching on child centred approaches to student motivation and emotional development (Bidwell 1965, pp.993, 995).

Bidwell's modification of the school as organisation is particularly significant for evaluation of teachers' work. For one, it defines the work of the teacher as something other than that of a rational bureaucrat executing a technical function. It must include both the impersonal judgement of accomplishments and the application of a personal relationship with the student to undermine resistance and facilitate learning (Bidwell 1965, p.993; Darling-Hammond 1997, pp.90-91). However, these two objectives are paradoxical and 'to perform adequately in his office the teacher is forced to violate the rules of performance' (Bidwell 1965, p.979. See also: Ingvarson 1995, p.23; Darling-Hammond 1997, pp.215-216). That is to say the teacher will often have to make a decision, adjust a procedure, bend a rule, or interpret a regulation based not on bureaucratic objectivity but with the specific interests or knowledge of the needs of a student. The personal and subjective will take precedence over the impersonal and objective, although the scale of this will likely be limited and indeed constrained by the teacher's awareness of the need for equity between students (Darling-Hammond 1997, pp.81, 94; Darling-Hammond 1988, p.61).

If one accepts Bidwell's argument, it is not sufficient to evaluate the technical aspect of teaching alone through ticking off functions from a duty statement. Evaluation must take into

account the ability of the teacher to overcome student resistance and demonstrate professional skill in the 'ability to handle day-to-day fluctuations in the response to instruction by individual students and collectively by the classroom group' (Bidwell 1965, pp.975, 1013. See also Connell 1985, p.129). Failure to do so is to revert to a superficial mechanistic evaluation form that is not only alien, but also irrelevant to the context. To go further, if these day to day fluctuations are significant to the process of education and yet the structural looseness of the organisation is such that the autonomy granted to the teacher makes him rather unobservable, then any failure to include a sufficient appreciation of the management of classroom fluctuations makes the evaluation irrelevant for any purpose other than the strictest bureaucratic measurement of the teacher's work. However, it is possible to be caught up in Bidwell's paradox again. He argues that the tendency of teachers to deal nurturingly and particularly with students tends to towards de-bureaucratisation in teacher-student interactions (Bidwell 1965, p.990). Thus a teacher's adequate performance from the students' perspective requires him to violate the bureaucratic principle of school organisation, and any evaluation of an effective teacher from an organisational perspective should be negative.

It is the question of the maintenance of professional discretion and colleague judgment against the criteria of standard procedures and the judgments of superiors. To the extent that professional discretion holds sway, coordination becomes difficult. To the extent that legal authority is emphasized, the advantages of educational expertness, that is, of variable procedures in response to variable students, are lost, and the alienation of professionally oriented teachers is more likely (Bidwell 1965, pp.1012-1013).

The staff appraisal systems introduced into schools over the past ten or so years have basically been transferred across from the commercial world and thus are characterised by a Weberian approach where performance is measured against a duty statement. Ingvarson and Chadbourne have noted that:

the appraisal cycle approach to teacher assessment is more bureaucratic than professional in its origin, application and intention. Members of long-established professions like medicine and law are not subjected to appraisal interviews...The public service, however, has adopted appraisal cycles with a vengeance and built human resource empires around the approach. Many models for teacher appraisal have been adapted from those used by the public service (Ingvarson & Chadbourne 1994, p.14).

Classroom discipline features significantly in such appraisal because the maintenance of control and discipline is a key element of the Weberian bureaucratic perspective (Perrow 1970, p.52). Whilst such staff appraisal schemes may superficially meet the needs of schools, they are incomplete from the perspective of Bidwell's analysis in that they do not penetrate or take into account the day-to-day fluctuations in the classroom or the affective bonds that the teacher establishes to further instruction. Consequently, the teacher is appraised as a functional labourer within the hierarchy adhering to the rules of procedure. When appraising officers think they are looking at teaching, they in fact can only look at the technical elements: lesson plans, record books, and mark schemes. In failing to look into the professional application of knowledge and judgement in the classroom, staff appraisal schemes look at only half the person, with lessened outcomes for both the individual and the organisation (Ball & Goodson 1985, p.11; Peterson & Chenoweth 1992, p.177).

Pusey's Dimensions of Organisation

Pusey (1976) builds on Bidwell's analysis by defining the *three dimensions of organisation* (Pusey 1976, p.31). The first dimension is the *formal structure* that defines the legal-rational authority of the school in a hierarchical framework as described by Weber. It consists of the hierarchy of the school, its positions and authority structures, statements of official rights, duties and responsibilities, and associated accountability systems. A second dimension that Pusey calls *technology* contains the sum of the cognitive and physical framework of education such as the syllabus, textbooks, classrooms, resources, and the timetable (Pusey 1976, p.32). A third dimension, entitled the *social system* takes into account the social dimension of an organisation, the interaction of personalities and the other psycho-social factors which are involved in determining the motivations and behaviours of its members (Pusey 1976, pp.26, 31-32). Pusey argues that one dimension alone is insufficient to describe the complexity of a school. Rather all three dimensions interact as independent variables within a dynamic system, with events in one generating responses in the other (Pusey 1976, p.18).

The school hierarchy will normally operate within the formal and technical spheres. The authority of the principal is derived by fiat of the bureaucracy or school board and structured in procedures and the hierarchy. At the same time, control is also exercised through specialised technical knowledge via the curriculum, designated syllabus documents, and the allocation of resources to classrooms and teachers. Teachers are part of the formal structure by way of their position in the school. However, the work of teaching involves the application of their technical knowledge and experience within the social environment of the classroom. As also described

by Bidwell, Pusey argues that teachers need to operate in a dynamic social environment where they are making judgements and decisions related to student interactions throughout the day (Pusey 1976, pp.22, 30).

Pusey's model suggests that the principal normally will make judgements within the formal and technical spheres, assessing teacher performance as a bureaucrat and technician, not for any other reason that these are the primary spheres of operation. The teachers, whilst they are part of the formal school hierarchy spend most of their time operating in the technical area of syllabuses and the delivery of learning. However, just as is argued by Bidwell, teachers also must develop affective bonds with students as they apply their personality to socially develop and not just technically educate. Consequently, successful teaching is more than just the delivery of the curriculum in a correct fashion. It also represents the judgement and experience of the teacher when to draw on his/her authority and expertise to achieve learning goals. From Pusey's point of view, this sets up an incompatibility between the bureaucratic/technical leadership system and the technical/social teaching system in essentially the same way as portrayed by Bidwell (Pusey 1976, pp.98-99).

However, Pusey's description of the response mechanisms triggered by accountability measures is particularly useful. For Pusey, when a criticism is initiated from the position of the principal (or his delegate) based on the formal and technical frameworks, this is not interpreted as such by the teacher who operates in the technical and social spheres. Because the teacher's personality and style are both inputs into the teaching process, the teacher transposes the criticism into his/her own technical-social frame of reference so that it becomes a criticism of the individual. In other words, because teaching involves the inclusion of oneself in the process of developing the affective bond with students required for good teaching, the teacher is unable to divorce the teaching from *self* (Pusey 1976, pp.97-98). Any evaluation system that does not engage the social dimension of teaching is likely to generate little more than personal resentment. Indeed, the need to protect ego will prompt teachers to minimise the opportunities for criticism on technical grounds and may generate social exclusion of the potentially punitive principal as part of the same protective process (Pusey 1976, pp.98-99).

Elmore's Loose Coupling

Elmore (2000) makes his position very plain in the first sentence of the foreword to his paper, stating that 'Public schools and school systems, as they are presently constituted, are simply not led in ways that enable them to respond to the increasing demands they face under standards-

based reform' (Elmore 2000, p.2). This statement is justified by a model of "loose coupling". It is argued that the "technical core" of education, the actual processes of grouping, instruction, assessment, and so on, reside in individual classrooms. The uncertainty inherent in this environment, flowing from the range of variables introduced by students, their individual needs, and changing teacher responses, means that the knowledge in the technical core 'cannot be clearly translated into reproducible behaviours, it requires a high degree of individual judgment, and it is not susceptible to reliable external evaluation' (Elmore 2000, pp. 5-6).

Elmore argues that school administrations have developed to "buffer" the weak technical core against inspection and disruption, and to ensure the maintenance of public confidence in what is a highly uncertain process. By this very function, administrations have moved their focus away from control of instruction out to the managerial and organisational functions that surround the instructional core; out to the line of defence, so to speak (Elmore 2000, p.6). Elmore seems to suggest that this shielding of teachers 'working in isolated classrooms under highly uncertain conditions', has become the prime function of the formal structure to such an extent that it has lost the understanding of that which it protects, such that even standards based reforms which would enhance teaching and learning, fail to penetrate through to the technical core. Along the way, they are bent and shaped into a form acceptable to the current organisational structure of schooling, and thus lose their efficacy (Elmore 2000, pp.4, 6)

The continuation of this state of affairs is underpinned by management hiring or selecting on the basis of administrative rather than instructional competence, namely those with the best "buffering" skills and able to maintain the level of confidence between the school and its constituency (Elmore 2000, p.7). It is also supported by the persistence of what Elmore calls "trait theories" of competence amongst educators and the public, which move the issue away from enhancing the system and developing mastery of professional knowledge to focus on personal qualities, which then become the basis for selection and retention of "better" people (Elmore 2000, pp. 8, 25). As a consequence:

The idea that people should acquire additional competencies over the course of their careers, that the organisation should systematically invest in the improvement of these competencies, or, more controversially, that people should be expected to meet higher expectations for competence over the course of their careers – these expectations don't exist, or exist weakly and idiosyncratically, in organisations that purport not to be able to manage their core functions (Elmore 2000, p.8).

For Elmore, school improvement based on standards reform, cannot exist in the current structure for a number of reasons. Because the administration fulfils its leadership (and ‘buffering’) function by virtue of managerial rather than instructional knowledge, it actually lacks an understanding of the technical core that it would ostensibly be trying to modify (Elmore 2000, pp.2, 23). Furthermore, the exposure of teaching to external scrutiny is such anathema to management that it simply could not allow such openness but would intrinsically modify the standards reform to coexist with the current state of affairs; effectively the management function is incapable of changing itself out of the “buffering” function. Lastly, management operations focus on control of functions within organisations, but the act of teaching cannot be controlled by virtue of its unpredictability, the reason for the existence of the external buffer in the first place. It rather requires the application of guidance, which, Elmore argues, presupposes a complete change of school leadership functions away from the current hierarchical managerial systems of control to a condition of distributed leadership where the skills and competencies of all in the organisation are harnessed more effectively towards a collectively agreed core function of the continual improvement of instruction:

Control implies that the controller knows exactly what the controller (if you will) should do, whereas guidance and direction imply some degree of shared expertise and some degree of difference in the level and kind of expertise among individuals. It is this problem of the distribution of knowledge required for large scale improvement that creates the imperative for the development of models of distributed leadership (Elmore 2000, p.14).

Loose coupling works against evaluation and the development of enhanced instructional practices because it encourages isolation and idiosyncratic explanations of good teaching based on personal qualities rather than knowledge. Improvement via the current school systems cannot occur because, in responding to initiatives from the formal level, teachers conform to organisational and not instructional objectives. Collaborative and collegial school events that ostensibly might be directed to reform can have the outcome of enhancing feelings of organisational well being, but are not translated into classroom teaching because they originate at the managerial level that is isolated from instruction, and by extension ignorant by lack of participation in it (Elmore 2000, p.17). Concurrently, under the influence of “buffering”, teachers know that they are protected from potentially threatening change and are able to continue inappropriate or outmoded practices. This is not to suggest any malicious or negative intent. Rather, Elmore would suggest that the current structure engenders a mentality of secure

complacency, where the sharing and opening to scrutiny of individual practice is to risk exposure and forfeiture of security.

Impact on Evaluation

Bidwell, Pusey, and Elmore all highlight the existence of a duality within schools that must be considered if the issue of evaluation of teachers' work is to be understood and procedures devised and applied in an effective manner. Whether it is termed "structural looseness", "loose coupling", or described as the function of elements in different dimensions, the operation of the managerial hierarchy and the teachers educating in the classrooms exist in a state of some tension. The differences between these interpretations tend to flow more from perspective and definition rather than philosophy.

Elmore's position derives from questions on the application of extended standards reform. Consequently, his analysis tends to be top-down and less directed to organisational factors operating between players within the school, except where structures and established processes impact on the reformist aim. Pusey, on the other hand, looks more within the school itself. For this reason, Pusey breaks down Elmore's "technical core" into both a technical domain of the curriculum, made up of curriculum, textbooks, classroom arrangements, setting and marking of tests etc, and the affective social domain of teacher-pupil interaction in the construction of learning (Pusey 1976, p.32). Bidwell has elements of both, referring to bureaucratization pressures on the school as well as the structural looseness required to give teachers autonomy to operate in accordance with their students' needs, and the affective bonds that must be established and which partially violate the apparent objectivity of the office (Bidwell 1965, pp. 976, 979).

Elmore helps explain the obstacles to evaluation of teachers' work on a broad level, and probably would dismiss any evaluation processes in place as either deficient, due to a lack of well defined standards, or else conducted by schools to provide a veneer of teacher improvement whilst not actually penetrating practice at all. Bidwell and Pusey identify a tension between their definition of the technical domain, including all of the observable aspects of teaching practice, and the professional side, where the process of interacting with students in such a way that learning and understanding, and intellectual and personal growth takes place (Pusey 1976, p.36). The technical aspect is easily measurable because the parts of it can be listed and deemed to be occurring or not occurring to a determined standard. In evaluation systems, these manifest as checks on aspects such as appropriate record keeping or lesson

planning (McLaughlin & Pfeiffer 1988, p.3). Once again, the maintenance of discipline figures prominently as an obvious manifestation of the teacher's skills (Shulman 1987, p.1). A technical perspective also lends itself to the one or two classroom visits that occur as part of staff appraisal systems. One can almost visualise the observing teacher noting down appropriate board use, good questioning technique and so on. For those who have experienced such observations, the "teaching to the test" feeling is very strong in the one being observed (McLaughlin & Pfeiffer 1988, p.4). In fact, it is not different from the observations we do of student teachers where we observe and check the technical behaviours of the pre-service candidate. The observing teacher is interested in what is observably done, not how the students are learning. In fact, the class remains the property of the instructing colleague teacher with whom the students share the affective bonds established over time. The pre-service teacher cannot hope to make that transition in most cases, and is likely to remain principally concerned with mastering the technical aspects; their personality may not manifest itself to the point where the real interactions that facilitate learning can take place. Notably, the same superficial process is accepted in staff appraisal schemes as being representative of the teacher's craft or competence.

The failure to penetrate and evaluate the actual learning event is probably due to a range of influences rather than one particular reason. Pusey describes a process of *withdrawal* where the tension between formal/technical position of the principal and associated power hierarchy causes the teachers to avoid disclosure of weakness or concern about the realities of teaching and to focus rather on self-protective behaviours. However, because of the close personal association of the teacher and his work, this also means the withdrawal of full disclosure of self in the work situation, and the relegation of the principal to his hierarchical role (Pusey 1976, p.99; Kauchak et al. 1985, p.33). In its worst manifestation, this may eventuate as strict adherence to formalistic behaviours that further obscure the classroom learning process (Pusey 1976, pp. 104-105). Ultimately such behaviour not only hides good teaching practices, and protects poor teaching, but also undermines the positive interaction between teachers and their willingness to take risks in pursuit of improved pedagogy (Pusey 1976, pp.99-100, 105).

Just because the appraisal of the technical side of teaching is seen as superficial does not mean that the *professional* position of teachers is automatic and inviolate. Bidwell and Pusey's analysis tell us where the process occurs organisationally. They do not suggest that effective learning will occur in each case. Rather, Bidwell suggests that the hierarchy deliberately accepts the structural looseness whilst the superficial control of the organisation overall

remains and general outcomes are generated (Bidwell 1965, pp.976, 1014). The autonomy of teachers in the classroom is accepted whilst the organisation focuses on ‘the reinforcement of the official component of the teacher’s role and the focussing of the commitments of teachers as professionals on the goals and policies of the system’ (Bidwell 1965, p. 1014). Pusey points out the futility of imposing further controls and similarly suggests that the social and technical dimensions of the teaching act protect it from bureaucratic controls, which indeed achieve greater levels of futility.

Pusey’s analysis provides an explanation of why teachers participate and accept staff appraisal systems when in fact these provide only a superficial view of their teaching, something that must be obvious to experienced teachers in particular. Teachers constantly feel vulnerable because of the psycho-social affective disclosure of their personality in classroom interactions. To protect themselves, teachers avoid any social affective engagement with the principal or assessors on the same plane, and consequently hide their personal attitudes, feelings, or thoughts from the perceived threatening formal level. This ensures there is no basis for revelation of

the *way individuals apply* the second dimension, that is, the technology and the cognitive and physical framework (the body of “educational science”, methods, the curriculum, the aims of education, etc.)...By virtue of the withdrawal game, discussion is always deflected away from the *process* (the third dimension or the “psycho-logic”) and directed instead to the “techno-logic”, which is elaborated in the hope that it will of itself take care of the process (Pusey, 1976. p.106).

Using this argument, staff appraisal restricted to the technical domain is instituted because it is as far as it is possible for the hierarchy to go. It fits the bureaucracy’s perception of duties and responsibilities and technical requirements. However, for teachers, to accept staff appraisal is to move discussion away from the *process* of education in the classroom to a purely technical level. They can feel safe because their personal input into the education process can be supplanted in favour of behaviours that can be addressed, are predictable, and limited. The price of self-protection and security is compliance with staff appraisal that depends on records of attendance at meetings, professional development statistics, and manifestations of classroom control. The adherence of school administration to staff appraisal is understandable in that it fulfils the bureaucratic accountability function (Kauchak et al. 1985, p.33). The assumption of

success in the function of teaching is through examination outcomes, and the progression of age cohorts through the system.

If Bidwell's structural looseness remains a characteristic of the system and Pusey is correct in identifying the gap between the formal/technical aspect of the hierarchy and the social/technical nature of the classroom, then the conditions do not exist for school administrations or educational bureaucracies to initiate an evaluation process that penetrates teachers' intimate pedagogy. Meanwhile, Elmore's analysis of "loose coupling" implies that the school administration in fact wishes to avoid penetrating the "technical core" of teaching itself, but would comply in response to external pressures and find a way to buffer the teachers, whilst adhering to the letter of the requirement instead of the spirit of the evaluation. This suggests that, whichever model is selected, if teaching is more than a technical process but is an application of professional skill judgement within a unique dimension, the verification of the quality of that professional process would have to rest with teachers, those who have both access to, and understanding of, all aspects of such performance (Stake 1989, p.17; Kleinhenz et al. 2001).

Lessons from Evaluation Practices

Targeting and Methodology

Darling-Hammond argues that the depersonalised factory model of teaching is aimed primarily at control and the transmission of a fact and rote based curriculum rather than learning for understanding (Darling-Hammond 1997, p.39). She defines recent trends in the United States for hiring less qualified teachers as a manifestation of bureaucratic dominance as the system does not expect teachers to be knowledgeable but to deliver pre-packaged materials and train students up to meet standardised test requirements. Shortfalls are responded to by greater control, increased bureaucratic surveillance, and further specification of learning requirements (Jones 1987, p.209; Tickle 2000, p.12). With this constrained structure, qualified and professional teachers are not necessary (Evans, 1999; Smyth et al. 1997, p.18). Furthermore, it perfectly justifies the use of test outcomes as measures of work output because the test becomes the pure object of teaching, rather than understanding and learning. The consequence of such a bureaucratically structured system, directed towards meeting central testing demands based often on public relations requirements, is that teachers have lost control of curriculum and learning (Darling-Hammond 1997, p.45; Evans 1999).

Darling-Hammond's analysis, of what she sees as the failed United States schooling system, highlights the fact that it is possible for teachers through inappropriate evaluation and questions of competence to be blamed for a whole range of ills within the system itself. The inability of bureaucracy to handle the complex and variable demands of education has generated a descending spiral of mandated outcomes and standardised testing in a bureaucratically directed attempt to portray improvement. Inadequacies in the process have generated even tighter controls and mandates, from testing down to prescribed competency levels and standardised teaching materials (Darling-Hammond 1988, p.56). The blaming of teachers becomes an almost inevitable consequence of it being impossible to find culprits within multi-tiered bureaucracies made up of complicated responsibility and decision chains. The teacher is easy to find and hold accountable (Darling-Hammond 1997, pp.52-53, 67).

Even where evaluation schemes are in existence beyond test scores, the literature is replete with criticisms of the application of evaluations, describing them as ritualistic or superficial (Shanker 1996, p.223; Shinkfield & Stufflebeam 1995, p.44; Murphy 1987, p.157). The process of classroom observation, a key component of many current evaluation processes, is particularly cited for criticism. Teachers argue that the selection of a microcosm of over a thousand hours of teaching at various grade levels at different times of the day, with various combinations of students in possibly different subjects and locations cannot in any way provide a representative sample (Bednall 1989, p.50; Stiggins & Duke 1988, p.87; Scriven 1981, p.251; Murphy 1987, p.159; Fullan & Hargreaves 1991, p.32). Teachers argue that learning is a construction of meaning in each class at each learning occasion, and that their abilities to foster learning across different students in one subject area or across different subject areas is inconsistent (Chadbourne & Ingvarson 1994, p.277; Schalock et al. 1993, pp.110, 123; Armiger 1981, p.297). Stodolsky similarly found that teacher behaviour varies with the subject matter and lesson format (Stodolsky 1988, p.133). Schalock's research showed that there was even no consistency in teacher's abilities to foster learning in students of differing abilities across subject areas (Schalock 1987, pp.59-60; Stake 1989, p.15). Darling-Hammond concludes that

it is impossible to effectively regulate methods of teaching. It is because public policy cannot act as an effective arbiter of such teaching decisions that it is important to ensure teachers' professional competence (Darling-Hammond 1997, p.221).

If the teacher's professional responsibility is to facilitate as much as possible the coming together of the optimum conditions for learning in any given class situation at any time, then

each class situation is unique because of the variables introduced by the teacher and the students as individuals, and their interaction to create a learning event. Schalock proposes a definition of teacher productivity as:

The contribution that a teacher is able to make to student learning by applying inputs that are relatively variable in the short run (pedagogy, time, classroom management, etc.) to inputs that are relatively fixed in the short run (inherent abilities, attitudes, school climate etc.) Given this definition, the contribution to student learning, rather than the level of student learning defines teacher productivity (Schalock 1987, p.61).

By extension, the quality of the teacher relates to his/her ability to facilitate such contribution and to apply knowledge in a timely and appropriate fashion (Ingvarson & Chadbourne 1994, p.11). However, a short term observation is not guaranteed to yield a specific event where the application of this skill occurs, particularly if the observed class becomes a staged event, aiming not to demonstrate skill but meet the anticipated preconditions of the observer (Darling & Hammond 1997, p.114; Fullan & Hargreaves 1991, p.33). Connell (1985) conversely argues that teacher's productivity is difficult to quantify because the product cannot be weighed or measured or sampled. Student scores may be pointed to, as in the United States, but this is not really the object; *learning* is the object and the test scores merely represent one measure of the process. Connell puts forward a substantial list of other immeasurable factors that range from discussing work with students, keeping records, and attending fetes, to planning and taking students on camps (Connell 1985, pp.71-72). This leaves very little real time for the teacher to engage in the sort of reflection and interaction that is going to enhance their practice (Holly 1989, p.110).

Ingvarson and Chadbourne (1994) criticise the appraisal cycle model currently in use in many Australian situations for being bureaucratic rather than professional in nature. Advocated more by employers and politicians, rather than teachers, appraisals have deteriorated to an annual ritual that has little to do with planning for professional development or career advancement (Ingvarson & Chadbourne 1994, pp.14, 31; Kleinhenz & Ingvarson 2004, p.35; Peterson & Chenoweth 1992, p.177). The appraisal cycle generally consists of a series of steps including an initial meeting, classroom observation and other data collection, an interview between the assessor and teacher, and any follow up or formal meeting with the principal. A more simple review process occurs in the second year (Ingvarson & Chadbourne 1994, pp.13-14). The

classroom visit attracts particular criticism from Scriven who condemns it on a number of counts:

First, the visit itself alters the teaching...Second the number of visits is too small to be an accurate sample from which to generalize, even if it were a random sample. Third, the visitors are typically not devoid of independent personal prejudices in favour of or against the teacher...Fourth, nothing that could be observed in the classroom (apart from the most bizarre special cases) can be used as a basis for an inference to any conclusion about the merit of teaching...Fifth, regardless of the fact that no observations of teaching style can legitimately be used as a basis for inference about the merit of the teaching, the visitor normally believes the contrary (Scriven 1981, p.252).

The appraisal cycle is unable to provide a clear indication of where the teacher needs to improve his/her teaching for the reasons outlined previously, although it may provide guidance on institutional behaviours. Indeed, Ingvarson and Chadbourne come to a similar conclusion to that voiced in the previous section, stating that appraisal schemes, particularly like those mandated above are aimed at:

exercising hierarchical surveillance, extending the dominance of a managerial culture, exercising political control and weeding out weak teachers. We also suspect that if these schemes were voluntary most teachers would refuse to participate because they lack the capacity to meet teachers' real needs (Ingvarson & Chadbourne 1994, p.15. See also: Connell 1985, pp.129-130; Gitlin & Smyth 1989, p.7).

Gitlin and Goldstein refer to this process as "dominant" in that, no matter how friendly it appears, where the evaluator is portrayed as expert who imposes standards on a teacher, the conduct of such an act of evaluation is inherently hierarchical, and thus threatening (Gitlin & Goldstein 1987, p.17; Gitlin & Smyth 1989, pp.7, 37).

The Issue of Competence

The question often motivating current evaluation events is generally "Is this teacher competent?" (Ingvarson & Chadbourne 1994, pp.31-32). The problem with this question is that nowhere in the literature is there a definition of what is a *competent*, or indeed an *incompetent*, teacher. Armiger (1981) argues that 'No conclusive research exists that determines what

characteristics of the teacher lead to learning on the part of the student. In fact there is no scientific basis to support the notion that one particular teaching style is superior to another' (Armiger 1981, p.297. See also Tickle 2000, pp.8-9). Bridges (1986) goes to some lengths to define incompetence, listing a series of failures, such as maintaining discipline, treating students properly, demonstrating mastery of the subject matter, and accepting advice from superiors. However, even whilst defining these areas, he counsels caution about a whole range of issues, namely:

- a. unsatisfactory performance may reflect shortcomings of the institution as well [also argued by Darling-Hammond (1997) who extends it to the whole public school system], and
- b. poor performance is rarely due to a single cause such as ability, but may reflect external factors as well (Bridges 1986, pp.7-8, 12).

Summative evaluations which highlight deficiencies meant to justify charges of teacher incompetence are mounted by a formal authority which may well have neglected its managerial function to provide or support formative evaluative measures for staff in its employ (Shanker, 1996, p.223). Bridges outlines various techniques, such as internal and inter-school transfer or relocation to administrative positions, which might protect the teacher but also deflect potential criticism of the administration and thus avoid unpleasantness (Bridges 1986, pp.25-26, 53; Armiger 1981, p.295). On this basis, taking steps against classroom teachers for students not achieving in standardised tests deals with the publicly perceived symptom of a shortfall, but it clearly avoids the cause. Nonetheless, the elimination of the incompetent or inadequate is a common purpose one associated with evaluation in United States literature. A rebuttal by the Miami-Dade County Public School Office of Professional Standards to criticism in the *Journal of Personnel Evaluation in Education* refers to revocation of teaching certificates, sex, and drugs, but makes no mention of the quality of teaching (Annunziata 1999). To extend the issue even further, Istance and Lowe (1991) question the use of the term "quality" because of the variability of the term based on perceptions and expectations (Istance & Lowe 1991, pp.39-41; Angus 1991, p.148; Soar et al. 1983, pp.243-244).

To refer back to Pusey's model, the bureaucracy deals with what it perceives as a personnel competency problem and makes sure it is seen to do so for its own purposes (Schon 1983, p.330). The formal level of the organisation cannot deal with the actual issue of teaching and learning because it lacks understanding of the particular second and third dimension processes at work (Darling-Hammond 1997, pp.74-75). The standardised test scores take on a life of their

own because they are mandated. No one bothers to investigate whether or not pupils from the particular school or class where the teacher has “failed” have greater understanding or developed qualities that will make them better citizens (Istance & Lowe 1991, p.44; Scriven 1981, pp.250-251; Soar et al. 1983, p.242). The strongest criticism of the “weeding out” approach to improving teacher quality is that from Ingvarson and Chadbourne who simply argue that the return for investing in removing the incompetent few is far less beneficial than concentrating on supporting the overwhelming majority of good teachers (Ingvarson & Chadbourne 1994, p.15; Duke & Stiggins 1986, p.15). Fullan and Hargreaves speculate that the level of irretrievable incompetents is no more than 2% or 3% and that:

schemes that implicate 100% of the staff in order to detect a small percentage of incompetents are a gross waste of time. Ironically the anxiety they generate can hold back the excellence of the many as they become reluctant to take risks for fear of punishment (Fullan & Hargreaves 1991, p.10. See also Duke & Stiggins, 1986, pp.14-15).

Theoretical Purposes of Evaluation

Purposes

Kleinhenz, Ingvarson and Chadbourne simply note that the purposes of evaluation are twofold: to guarantee teaching quality, and to improve teachers’ work (2002, p.2). However, a tension is created between the two that has undermined evaluation systems and made it ‘difficult to “get it right”’ (Lokan & MacKenzie 1989 p.2). Stake similarly points out that whilst ‘these purposes regularly coexist, they get in each others way’ (Stake 1989, p.13). In his review of teacher education in New South Wales, Ramsey (2000) defined the issue as one of quality assurance of teachers’ work versus accountability. For quality assurance

Self-improvement and self-regulation of individual units, parts of organisations or the organisation as a whole is the main aim. This process of continuous improvement is based on formative judgements about how to meet stated objectives and goals. Accountability, more commonly, arises from the efforts of external bodies or whole systems to measure outcomes, and to hold managers accountable for processes, programs and outcome levels achieved. Benchmarking and comparison with other similar enterprises in terms of productivity and quality are core issues (Ramsey 2000, p.120).

The tension flows from the fact that teachers are not going to be willing to expose their weaknesses to those who will determine their tenure or promotion (Ramsey 2000, p.120). This is not a new argument. Beare not only makes a distinction between the types of evaluation but also considers the role of both the audience and the assessor as significant (See Figure 1 in Beare 1989, p.15). Thus, broad statements such as “teacher quality” are inadequate. Bednall further comments that teacher evaluation cannot be generic in application: ‘They must be site specific and a product of the school context in which they will operate’ (Bednall 1989, p48; Holly, 1989, p.112). If this is the case, not only do multiple purposes exist, but also the application of evaluation to each environment may require a specific form so that no generic solution is available, nor perhaps even possible. Furthermore Stake and others caution that it is insufficient to evaluate just teaching; the evaluation environment must also examine institutional goals, the administrative organisation and operations, as well as what is taught (Stake 1989, p.13; Scriven 1981, p.245; Schalock 1987, p.60; Schalock et al. 1993, pp.109, 125; Darling-Hammond 1997, pp.11-12).

Generally, teacher evaluation processes have been summative and accountability directed, making them the source of widespread dissatisfaction, anger and fear (Millman 1981, pp.279-80). Mertler (1999) argues that the improvement of instructional practice (formative) has been under-utilised in the United States as opposed to evaluation for selection and retention, and that the latter has been relied upon as the sole source of information about a teacher. In Victoria up to the late 1980’s evaluation of teachers in government schools occurred only in response to applications for permanency or promotion. Both were summative in form and accountability directed (Ingvarson 1989, p.140). Unfortunately, in such evaluations the teacher ‘will be at pains to hide from the person or persons carrying out the review any inadequacy in his or her teaching practice, and to put before the assessor all the strong points in his or her performance’ (Beare 1989, p.13). Handy notes that individuals are unlikely to admit to deficiencies in any situation where it will provide data for future assignments; performance in a current position is not necessarily the best basis for assessing capabilities in a promotion situation (for example a very good teacher may not be very competent at administrative tasks); and that the effectiveness of traditional approaches has been undermined ‘in large part, because the superior is expected to be, at the same time, judge and counsellor’ (Handy 1981, pp.253-254; Popham 1988, p.270). Kleinhenz and Ingvarson see this issue as a key weakness of the experienced teacher with responsibility procedure in Victoria where both parties feel the ‘need to create and preserve certain barriers to in-depth conversations about practice’ (Kleinhenz & Ingvarson

2004, p.39). This is followed by a clear statement that this recent evaluation event in Australia was discredited by a number of serious flaws:

The evaluation provides only a façade of accountability, holds little capacity to understand practice, fails to give credible recognition to teachers whose work is highly accomplished, and offers little information for teachers to improve their practice (Kleinhenz & Ingvarson 2004, p.40).

Implementation of any evaluation process should be accompanied by clear objectives and responsibilities. In the case of tenure and promotion, it is the school hierarchy or bureaucracy that initiates the evaluation process. Referring back to the organisational analysis earlier in this chapter, it is an evaluation from the rational-formal dimension; top-down and accountability directed. As argued previously, the inability of this level to penetrate actual classroom processes in a meaningful way means that it must make its judgements from only the observable technical and mechanical processes of teaching (Murphy 1987, p.158; Loup et al. 1996, p.216). Namely, an observer may comment on classroom discipline and the quality of the notes on the board, but is incapable of making a judgement of the quality of learning in the absence of exposure to the normal classroom environment over time (Shulman 1987, pp.6-7; Darling-Hammond 1997, p.75). Consequently, it is not fair to include any judgement for tenure or promotion with evaluation that is directed to the improvement of student learning or teacher practices directly related to the core aim of the education process. The converse applies as well (France 1989, p.75; Stake 1989, p.13).

Kleinhenz, Ingvarson and Chadbourne note that ‘Schools cannot “implement” teacher evaluation systems if they do not know why they are doing so’ (Kleinhenz et al. 2002, p5. See also Cumming & Jasman 2003, p.16). There is an issue here with use of the term *schools*. If it refers to the formal organisation, the implementation will automatically represent the aims of the hierarchy and be directed towards the technical side of the evaluation process. In this case, school should clearly state that this particular evaluation is aimed at measuring the duties and responsibilities of the teacher as employee. However, if the statement refers to the school as a learning community, then, indeed, the focus needs to be on learning and development beyond the formal structure (Stenhouse 1975, p.166). In this case it might be better if the quote had commenced with *Teachers*. The problem is that teachers as a whole have so far not initiated the management of the quality of their work as professionals (Darling-Hammond 1997, p.213).

The Standards Model

Ingvarson and Chadbourne, consider that 'most teachers want to do a better job and have a natural desire to become more competent. Teacher evaluation schemes should reinforce this desire' (Ingvarson & Chadbourne 1994, p.19. See also Cumming & Jasman 2003, p.16). It has been argued that no fundamental change in education is possible unless it progresses through the teacher into the classroom (Stenhouse 1975, p.25). Toward this objective, Ingvarson and Chadbourne have long advocated a career development model for enhancing the positive aspects of teaching. This model consists of three characteristics: career stages, standards, and evaluation (Ingvarson & Chadbourne, 1994, p.16). They propose that teachers' careers switch from horizontal mobility where they just change schools to a vertical model where status and remuneration are based on objectively judged competence. Standards are to be based on what teachers should know and be able to do, such as those developed by the National Board for Professional Teaching Standards (NBPTS) in the United States. The NBPTS model is a summative one centred on satisfying a select and objective, as well as subject specific, teaching panel of the candidate's worth and tangible demonstrated skills through a comprehensive portfolio. Experience allows examination of not just competencies but also the pedagogical reasoning behind a teacher's work (Ingvarson & Chadbourne 1994, p.33). Assessment by colleagues from the same subject area allows analysis of content specific expertise. The bases for this are the findings of Stodolsky and others that teacher behaviours and approaches vary systematically with subject matter (Stodolsky 1988, p.133; Ingvarson & Chadbourne 1994, p.33).

The argument is that such a standards body not government bureaucracies would be the determiner of what is best practice, in the same way that health departments administer the health system but the medical profession sets standards (Ingvarson & Chadbourne, 1994 p.24). Such developmental standards are not constrained but assume that 'teaching is a job you can get better at and never stop getting better at. They assume that there is always room for a teacher to increase his or her knowledge and skill' (Ingvarson & Chadbourne 1994, p.29). Because it is voluntary and does not affect tenure or promotion directly, teachers are encouraged to participate for their own professional purposes and to concretely manifest the adage that teachers want to teach better as has been claimed. The voluntary characteristic would act as a force for encouragement instead of coercion. As teachers who achieve the standards derive greater success from enhanced skills and knowledge, and self-confidence in their teaching, the subsequent rewards that might flow from promotion are thus one positive by-product.

More than anything else, the establishment of an independent and respected institution which sets professional teaching standards would assist the development of teachers and the cause of effective evaluation in that it transcends employers, authorities, and teacher training institutions and sets the standard for accomplished practice (Darling-Hammond 1988, p.59). Meanwhile, the academic rigour of the National Board for Professional Teaching Standards process was verified by only about half of the applicants being admitted in 2000 (ACER 2000, p.8). As Ingvarson and Chadbourne note, it could assume the function of registration (which in the Tasmanian sense certainly is once again not professionally but governmentally mandated) and award certification of teachers for advanced positions. The selection of a candidate for a position would remain the prerogative of the employer, but there would be no doubt who was the better qualified teacher (Ingvarson & Chadbourne 1994, p.24).

The criticism of the standards movement is that it is knowledge-based and restricted to the technical area of the teaching process. Inputs are derived from the social level of the classroom where professional judgement is applied through a teacher's portfolio and videos of class teaching. As such it reflects only a general or isolated description of the classroom environment where a teacher applies a large number of judgements in any one lesson. As the panel reviewing a teachers input is completely anonymous and objective, it cannot identify directly with the environment or personalities involved in a teacher's work. Nevertheless, it does fulfil a major requirement in defining standards and competencies for teacher knowledge. The fact that it is removed from the organisational context and administered by teachers with pedagogical and subject specific knowledge certainly moves evaluation beyond its current situation and establishes a valuable component of the end result being sought.

Professional standards certification is summative in that a judgement is made on the teacher's professional standing based on demonstration of knowledge and expertise measured against an agreed standard. However, it concurrently assumes that ongoing formative development of teachers is enhanced by certification. The process is meant to be liberating and continuous rather than an end point. It differs from hierarchical judgements for tenure and promotion that yield decisions about employment. If this distinction is recognised then not only are both fear and confusion diminished, but the cause of evaluation advanced. (Cumming & Jasman 2003, p.19).

The Duties-Based Approach

Scriven contends that a duties based approach is more effective than the reflective National Board for Professional Teaching Standards model as the latter is ‘only a measure of how well someone *could* do, not of how well they *do*’ (Scriven 1994, p.84). However, his major criticism is directed to competency-based models which are considered invalid as they are based more on judgements of style: the example is cited of lesson plans, for which there is no pedagogical justification. More significantly, Scriven highlights the fact that teachers have duties that do not manifest in empirical approaches, such as caring for the children in emergencies (Scriven 1994, pp.83-84). What is not clearly outlined is the form or structure of the duties approach although there is a requirement for data gathering on a broad front, including student feedback, a teacher portfolio, and possible subject specific input. It is ‘simply the evaluation of teachers against the criteria specified in a comprehensive list of teacher duties, modified as appropriate to fit a particular site, district, or state’ (Scriven 1994, p.71).

The strength of Scriven’s duties-based model is that it does take place in the school and is likely to be expandable to a whole range of teacher duties beyond pedagogy, where Ingvarson and Chadbourne are more focussed. As such it does allow a better focus on community and school needs and their adjustment to changed requirements. The difficulty is that a duties-based approach lends itself to a checklist or rating scale outcome. It has also yielded hybrids, such as the Danielson (1996) model. Danielson claims that her work ‘represents all aspects of a teacher’s responsibilities that are reflected in daily work’ (Danielson1996, p.14). It lists as *Elements* many of the duties listed by Scriven, but omits duties such the management of emergencies, which is a fairly important part of the duty of care responsibility. Furthermore, the style criticisms made by Scriven are transferred to the Level of Performance categories. Scriven would argue that the Element of Learning Activities defined as *follow an organized progression and....reflect recent professional research* is a judgement by the observer and invalid. As Scriven states:

If it is something that we find or think to be - in the basis of research or experience or a theory - a good or successful approach, such as acting as a resource instead of a source (as a discussion initiator instead of a lecturer), or using a multimedia approach when one is available, then it cannot be used as a criterion of merit (Scriven 1994, p.83).

By extension, the ultimate measure is whether the children learn effectively, which is why Scriven advocates consulting the students (Scriven 1994, p.87). To be fair to Danielson, she does not assert that her work is for summative evaluation which Scriven advocates, but mentions only its use with mentors and coaches suggesting that it is more a formative tool for peer assistance (Danielson 1996, p.20). Scriven does not deny the value of such activities providing that one acknowledges that they are directed towards style, and that consequent decisions on merit lack validity (Scriven 1994, p.85; Soar et al. 1983, p.243).

Peer and Self Appraisal

Within the formative area, there has been support for peer appraisal, working with one's fellow teachers to help improve teaching and overcome difficulties. The cooperation of teachers is to be welcomed at any time and indeed peer-based efforts may yield some benefits. However, peer appraisal can be criticised on a number of counts. Unless peers are directed to cooperate in such a system, which by the very fact of *direction* makes it a manifestation of hierarchical control, it is probable that teachers will select like-minded colleagues who may not challenge existing practices or could be uncritically generous in their assessments. Scriven most strongly criticises peer appraisal as being typically style assessment and thus invalid (Scriven 1994, p.85; 1989, p.115; Peterson & Chenoweth 1997, p.179). There is also no compulsion to change and an "agree to disagree" compromise may eventuate. Unless one or more of the peers is particularly motivated and willing to do the research, peer approaches also are unlikely to stay abreast of best practice. As Lortie found, peer dialogue is generally related to picking up "tricks of the trade" rather than broader pedagogical concepts (Lortie 1975, p.7). Finally, the proliferation of peer groups each with its own specific agenda is unlikely to provide the co-ordination necessary for either the enhancement of teaching practice across a school or the effective management of change. Any peer based system would also need to break through social issues that threaten to undermine it, such as the teacher in Knoblock and Goldstein's research who:

actively ostracized by the other teachers on her grade level because of the type of classroom she was conducting, stated that it made her feel better when she realized the others were really threatened by her competence and that it wasn't just that she was a poor teacher (Knoblock & Goldstein 1971, p.9. See also Darling-Hammond et al. 1986, p.229).

Ideally all teachers should be reflective practitioners, the sort of teacher who Schon says:

sees a child's difficulty in learning to read not as a defect in the child but as a defect "of his own instruction". So he must find a way of explaining what is bothering the pupil. He must do a piece of experimental research, then and there, in the classroom. And because the child's difficulties may be unique, the teacher cannot assume that his repertoire of explanations will suffice, even though they are "at the tongue's end". He must be ready to invent new methods and must "endeavour to develop in himself the ability of discovering them" (Schon 1983, p.66).

However, the lack of widespread teacher engagement with evaluation to date suggests that few teachers may be reflective in this form, even if they are what many may term "good" teachers. Jones notes that 'Teachers are not, by nature of their choice of profession, the most adventurous, risk-taking and daring of people. The daily round, the common task, take their toll' (Jones 1987, p.199). Alternatively, they may have settled into the complacency flowing from becoming too skilful at their techniques:

Many practitioners, locked into a view of themselves as technical experts, find nothing in the world of practice to occasion reflection. They have become too skilful at techniques of selective inattention, junk categories, and situational control, techniques which they use to preserve the constancy of their knowledge-in-practice. For them, uncertainty is a threat; its admission a sign of weakness (Schon 1983, p.69).

Whilst an element of self-evaluation may be present and desirable in other evaluation approaches, an individual teacher is unlikely to be capable of practising self-evaluation effectively on a continuing basis for a number of reasons. Firstly, we are unable to effectively criticise ourselves because we cannot by definition be objective (Trethowan 1987, p.70; Stodolsky 1988, p.135). Others have to help teachers appreciate their approaches and find better ways (Schon 1983, p.333). As Fullan and Hargreaves state: 'Our own experience is partial. Deeper reflection requires other eyes, other perspectives as well as our own' (Fullan & Hargreaves 1991, p.68). Bednall's article on *Self Appraisal: A Philosophical Framework* introduces a "participating colleague" to moderate, help devise an instrument, and provide some praise and recognition (Bednall 1989, pp.51-52). Secondly, if teachers are not aware of what transpires in the next classroom or those beyond, they are unlikely to have a valid frame of reference for examination of their own work (Stenhouse 1975, p.159). Many teachers 'whose classrooms reveal even more profound avoidance believe they are doing good work'

(Powell et al. 1985, p.115. See also: Fullan & Hargreaves 1991, p.39; McLaughlin & Pfeiffer 1988, p.64). Thirdly, as with peer appraisal, individuals alone are unable to remain conversant with all literature about issues and practices in education (Griffin 1986, p.75). Fourthly, solitary self-reflection does not necessarily translate into improvements across the profession and the consequent enhancement of education in the broader society (Ingvarson and Chadbourne 1994, pp.26-27; Fullan and Hargreaves 1991, p.67).

Schon's reflective practitioner is not solitary and isolated, but rather one who eschews certainty and welcomes unexpected experiences, be they 'pleasing and promising or unwanted' (Schon 1983, p.56). Such teachers undertake further academic study, or become action researchers, to try and resolve issues of concern about teaching, including their own practice. Such dynamic teachers may draw in others in an ongoing peer study group that has evaluative as well as professional development functions (Viadero, 2002; Distad et al. 2000). At the same time, they likely will be institutionally confrontational as they clearly will perceive the conflict between the imposition of bureaucratic control and the right of teachers to practice as an autonomous professional (Schon 1983, p.337).

Discussion of Approaches

Summative and formative evaluation are equally important, and researchers see a need for them to be applied in the appropriate circumstance, for the correct reason, in an open system (Ingvarson & Chadbourne 1994, p.31). Scriven criticises formative evaluation alone as not actually committing itself and thus being open ended to the extent that it is not really evaluative (Scriven 1991, p.29). This is the criticism that is often levelled at peer appraisal systems; they can become purposeless as they focus on minor aspects of teaching without coming to any conclusion. Handy argues even further that each specific function - data collection, performance assessment for reward, individual feedback on performance and personal strengths and weaknesses, and goal setting - should be the subject of different procedures by different people at different times (Handy 1981, pp.253-254). Darling-Hammond and her colleagues interestingly come at the problem from the other direction arguing that the data outcomes for each evaluation process should be such as to be useful rather than confusing:

In general, teacher evaluation processes most suited to accountability purposes must be capable of yielding fairly objective, standardised, and externally defensible information about teacher performance. Evaluation processes useful for improvement objectives must yield rich, descriptive information that illuminates sources of difficulty as well as viable courses for

change. Teacher evaluation methods designed to inform organisational decisions must be hierarchically administered and controlled to ensure credibility and uniformity. Evaluation methods designed to assist decision making about individuals must consider the context in which individual performance occurs to ensure appropriateness and sufficiency of data (Darling-Hammond, et al. 1986, pp.224-225).

No single approach to evaluation seems to meet all requirements (Tickle 2000, pp.10, 19). The professional standards model, supported by Ingvarson and Chadbourne, certainly covers the higher order elements of teaching, and demands a demonstrated standard of knowledge, and intellectual rigour and commitment to achieve accreditation (Ingvarson 1998b, p.1028; ACE 2000; Ingvarson 2001; Kleinhenz et al. 2002). However, it does not deal with the day-to-day processes of teaching and some of the duties and activities in which teachers engage in the affective domain, the counselling and support that helps students through a critical time in their lives (Hargreaves 1997b, pp.105-106). This approach is also very academic and requires a strong personal commitment to complete the assessed portfolios. Furthermore, whilst it may be comprehensive and encourage reflection on and analysis of pedagogy, it does appear slightly impersonalised as the submission is sent to a panel remote from the classroom and schoolyard. The National Board for Professional Teaching Standards has also been criticised for being 'aimed at identifying and rewarding the elite of the American teaching profession' (ACE 2000, p.8). Furthermore, research by Schalock has shown that it is not just knowledge and skills that make a teacher effective, but the timely and appropriate application of both (Schalock 1987, p.62). Whilst it contains routine elements, teaching remains a non-routine task that involves *judgement*. Such judgement is the response to the potentially unfamiliar in each class where the teacher must rely 'upon a residue of something we do not understand at all well - experience, judgement, knack, wisdom, intuition' (Perrow 1970, p.76). Tickle goes further, stating that 'at the core of the job, it remains complex to the point of being undoable' (Tickle 2000, p.12). The best place to appreciate this event would be within the school (Schalock 1987, p.62). Scriven's (1994) duties approach certainly allows coverage of most functions in a teacher's life and is available for modification based on local needs. However, it does not probe into the areas covered by the standards model, attempting to verify and accredit the problem solving and teaching style variations that a teacher applies to help students learn.

The Issue of Change

Change is relevant to the discussion of teacher evaluation both because we live within a rapidly changing society, and because evaluation, of itself, implies a change through enhanced performance. Social change implies not just the altered social, intellectual, technological, and often emotional, profile and needs of students, but also world of work. As Ramsey found:

There are now few occupations, vocations or professions that have not had to respond to the collapse of many of the certainties once underpinning their work. In the process they have re-defined how they conduct themselves and the role they have in our society (Ramsey 2000, p.19).

For education systems to respond effectively to new conditions, all aspects of it need to move forward together (Stenhouse 1975, pp.25, 143).

The Queensland New Basics Technical Paper notes sociologist claims that information technology globalisation is doubling human knowledge each decade, and that curriculum development cycles of fifteen years may be far too long (Education Queensland 2000, p.29). On the social plane, marriages and birth rates decline whilst divorces increase. Many jobs have disappeared and new ones are being constantly created (Education Queensland 2000, p.12). Schools have had to respond to socially mandated equity programmes, such as mainstreaming of the disabled, and multicultural agendas, such as the inclusion of non-English speaking students in classrooms, all of which have required additional input from teachers (Istance and Lowe 1991, p.27; Scott et al. 2001; Commonwealth of Australia 1998, pp.44-45; Elmore 2000, p.2; Tickle 2000, pp.6-7). Cries for “returns to the basics” are coupled with concerns that structural and organisational changes have not yielded the benefits expected, whilst claims abound of employers complaining about graduate skill levels (Istance & Lowe 1991 p.25).

A common response of both Government and the bureaucracy is to amend requirements and assessment procedures as the educational systems of society try to equip students with skills and knowledge to better meet today’s needs and those of the predicted future. In other words, they respond to the change by making superficial alterations within the acceptable bureaucratic guidelines (Perrow 1970, p.64; Elmore 2000, p.6). Yet, at the same time, changes are imperative in schools and classrooms where traditional structures and methods become increasingly irrelevant (Education Queensland 2000, pp.12, 20). Linda Darling-Hammond goes even further, suggesting that current education systems are failing because their rigid and bureaucratic structures were never meant to confront the demands now being made (Darling-

Hammond 1997, pp.xi, 27; Jones 1987, p.203). Elmore expresses a similar sentiment in stating that ‘Schools are being asked...to do things they are largely unequipped to do. School leaders are being asked to assume responsibilities they are largely unequipped to assume’ (Elmore 2000, p.2). Against this background, calls for open government and institutions have resulted in schools becoming increasingly porous and permeable institutions that can no longer conceal what they do (Freeland 1991, pp.79-81; Hargreaves 1997, p.86).

Schools no longer have a monopoly on learning and represent just one element of the development of youth (Griffin 1986, p89). In a high-speed technological society, where travel is cheaper and faster and information readily accessible over the Internet, some students will have greater knowledge and experience than those teaching them. Computer skills are perhaps the obvious example (Hargreaves 1997a, pp.5-6; Jones 1987, p.210; Istance & Lowe 1991, p.34). The pace of change demands new types of workers:

The speed and ubiquity of technological innovation postulate a demand for workers capable of adapting flexibly to new demands by acquiring knowledge and skills that facilitate occupational change, and by acquiring the breadth of expertise needed to discharge a variety of tasks in work organised into teams (Istance and Lowe 1991, p29).

Hargreaves is happy to argue that schools need to develop students as knowledge workers to equip them to operate in an information society (Hargreaves 1997a, p.9). He expands by saying that there is no longer any room in society for dominant and excessively didactic methods of teaching, or for placing content and mechanical process above understanding. Demands for educational change flow from the increasing alienation engendered by traditional practices which are perceived by consumers as lacking in value and which do not motivate students (Education Queensland 2000, pp.20, 31). Students are no longer seen as blank slates, but as individuals whose unique learning styles must be accommodated, and whose cultural backgrounds should also be taken into account (Education Queensland 2000, pp.12, 19, 29). Added to this is the fact that schools and teachers are expected to be aware of, and accommodate, external emotional and social impacts on student’s lives (Istance & Lowe 1991, p44; Hargreaves 1997a, pp.7-8).

These factors all require input from teachers beyond the basic “chalk and talk” function in the development of students as whole people. Hargreaves notes that teachers consequently have to:

reinvent their sense of professionalism so that it does not place them above or set them apart from parents and the wider public, but gives them the courage and confidence to engage openly and authoritatively with others who also have the children's interests at heart (and even those who don't) (Hargreaves 1997a, p.11).

He further argues that the emotional element of good teaching, coupled with the extant and further desired permeability of schools means that schools will need to move beyond traditional hierarchal organisation, and more recently managerialism, into a more open and mutually supportive relationship with the community (Hargreaves 1997a, pp.17-18; Darling-Hammond 1997, p.31; Trethowan 1987, p.62). In this case, schools and teachers will need to communicate more openly with their constituency, and school reports will need to go beyond being 'insurance policies against potential parental criticism' (Hargreaves 1997a, p.20).

The perceived inadequacies of the education system yield bold new visions and changed assessment methods: New Basics, and Essential Learnings (Department of Education Tasmania, 2002). These visions of meeting the needs of the future flow from the top down, from visionary researchers, bureaucrats, or politicians, who become increasingly eager to demonstrate that they have the solution to the problems afflicting literacy and numeracy levels or curricular relevancy (Perrow 1970, p.64; Elliott 1998, pp.17-18). The problem is that such bold visions are often devoid of the practical procedures required for their implementation into the classroom (Darling-Hammond 1997, p.5). Much is left up to *professional judgement within approved guidelines* so that 'the rhetoric of reform is largely a paper and pencil matter' (Cohen and Barnes 1993 p.245). As Cohen and Barnes note in looking back at the poor record of change in the USA: 'One thing that is missing in most reform proposals is a sense of what the new teaching might look like and what teachers would have to know and do in order to carry it off' (Cohen & Barnes 1993, p.242. See also Darling-Hammond 1997, p.216).

Despite frequent loud claims of confidence in the professional standards of the teachers, bold visions and initiatives often do not succeed (Fullan & Hargreaves 1991, pp.13, 98-99; Dimmock 1995, p.274; Darling-Hammond 1997, p.xiv; Istance & Lowe 1991, p.147). Fullan points out that 'Mandates are not sufficient and the more you try to specify them the more narrow the goals and means become. Teachers are not technicians' (Fullan 1997, p.29. See also Fullan & Hargreaves 1991, p.102). As evidence of the lack of success, Istance and Lowe refer to the preponderance of teacher talk in the classroom and a remarkable stability of education

practices across a number of countries despite a whole range of reforms (Istance & Lowe 1991, p.25; Angus 1991, pp.152-153; Trethowan 1997, p.68; Darling-Hammond 1997, p.214). Whether the hindering factor is the teachers themselves, or the outcome of Elmore's "loose coupling", the effect remains the same.

Fullan and Hargreaves stress the need for reform to be realistic, achievable within time lines, avoid alienating teachers, and to focus on motivating teachers to change their practices (Fullan & Hargreaves 1991, pp.13, 16-17, 20). Darling-Hammond, Wise, and Pease see effective change as requiring 'knowledge that a course of action is the correct one and a sense of empowerment of efficacy' to ensure it occurs (Darling-Hammond et al. 1986, p.237; Darling-Hammond, 1997, p.217). This is acknowledged in the Queensland New Basics which states that:

Adjustments in curriculum and assessment in and of themselves cannot generate changed outcomes without a focus on classroom interaction, where students and teachers select, shape and renegotiate contents, skill and knowledge, and their assessment, in idiosyncratic ways (Education Queensland 2000, p.33).

Fullan asserts that changes of value require new beliefs and behaviours and not just skills, whilst Barnes and Cohen note that 'The teaching that reformers seem to envision would thus require vast changes in what most teachers know and believe...they would have to learn to teach differently' (Barnes & Cohen p.246. See also: Fullan & Hargreaves 1991, pp.13, 25; Fullan 1997, p.43; Tickle 2000, p.11; Jones, 1987, p.202).

The Tasmanian Essential Learnings Framework 1 (2002) document contained numerous statements about what educators do. They:

View curriculum as being based on key ideas or questions (p.42).

Negotiate with learners about learning goals, activities, issues and contexts (p.43).

Make explicit the connections between different aspects of learning (p.44).

Maintain a sense of humour (p.45).

Involve learners in goal setting (p.47).

Explicitly teach thinking and learning strategies (p.49).

Design learning sequences that explicitly support the transfer of learning to new problems and settings (p.49).

However, these slogans remain exactly that until they are translated or incorporated into instruction and learning tasks in the classroom by teachers who may have to change how they feel and what they do. At the very least, any change must capture sufficient teacher commitment to have any chance of success. Yet this aspect of effecting change is generally ignored (Fullan 1997, p.30; Fullan & Hargreaves 1991, p.13; Angus 1991, p.151; Ingvarson, 1995 p.24; Darling-Hammond 1997, pp.9-11). Howe argued that, in the United States, the prescriptions of how and what to do, and demands of standardised testing, undermined change by constraining the teacher. He asserted that teachers needed to be given more freedom and latitude if improvement was to be successful. He noted that ‘Leadership which taps the creativity of those who are at the centre of the project, whatever it is, will always do better than leadership that uses its authority’ (Howe 1994. See also: Darling-Hammond 1997, p.214; Angus 1991, p.152; Evans 1999). Ramsey reported a similar finding in New South Wales:

Teachers will only successfully prepare tomorrow’s ‘knowledge workers’ if in their professional practice they too are “knowledge workers”. This is not to say that good teachers do not adapt; they do, but they are constrained by systems seemingly more focused on compliance rather than emphasising and valuing the professional capacity and responsibility of the teacher (Ramsey 2000, p.24).

Movement away from the isolated classroom where content learning is the priority to a multidisciplinary framework where integrated tasks lead students to discover unique solutions to problems is not possible unless teachers make it so. Teachers need to be capable, willing, motivated and supported to make the change (Fullan & Hargreaves 1991, pp.13-14, 27; Hargreaves 1997a, p.30; Freeland 1991, p.82; Stenhouse 1975, p.25). If the changes that are designed to resolve a problem are presented to those responsible for the operation of the system without an examination of the factors that contributed to previous shortfalls, and without capturing teacher enthusiasm for resolution, the generation of systemic and consequent failure seems a natural consequence (Jones 1987, p.202). The initiative will confront inertia. As Hedberg suggests, there is little inducement to change if established procedures work perfectly well (Hedberg, 1981 p.5). Ramsey identified this inertia in his review.

It remains a fact, however, that in too many instances students, especially in the upper primary and secondary years, describe the teaching they experience as consisting of such practices as copying blackboard notes, taking dictated notes, writing up summaries from textbooks, watching large

numbers of videos and doing work which seems more about occupying time than purposeful learning (Ramsey 2000, p.25).

Perrow goes beyond the teachers to suggest that bureaucratic structures are able to accommodate a fair amount of superficial change without actually changing themselves, whilst arguing of course, that the transformation has been profound (Perrow 1970, p.64; Stenhouse 1975, p.169; Elliott, 1998 pp.24-25). This accords with Elmore's more recent concept of "loose coupling", but in this case a change does occur through the manipulation of the innovation into existing frameworks, concurrently becoming no change at all (Elmore 2000, p.4). Darling-Hammond supports this view stating that superficial reform is inevitable due to the fact that 'most school systems do not know how to manage the intensive engagement with ideas and people required to enact broad policy change' (Darling-Hammond 1997, p.217). In reviewing the need for changes to education in society, a similar view appears to have motivated Ramsey to comment in his review that 'The achievements of our best teachers point to the salient fact that change in the classroom comes not from imposed priorities and bureaucratic regulation, but is drawn from the deep well of the teacher's professionalism' (Ramsey 2000, p.25). If such professionalism is ignored or threatened then the prospects for success must diminish. Ramsey put it most succinctly when he wrote that 'Significant issues arise for students in those instances where the quality of teaching lags behind the quality of the curriculum; the effect on student learning is less damaging where the reverse applies' (Ramsey 2000, p.13).

Fullan and Hargreaves identify a requirement for teachers to move away from isolation and tap into each other's ideas. Consequently, evaluation should be directed to opening up the capacity of the majority of teachers to work together to fully utilise their expertise in pursuit of better learning, independent of curricular change and on a continuing basis (Fullan & Hargreaves 1991, pp.10-11, 38-39; Fullan 1997, p.39). Terms such as teaching from other perspectives, reflection, and a dialogic orientation towards teaching and learning are increasingly common in the literature (Cohen & Barnes 1993, p.246; Gitlin & Smyth 1989, p.61). The isolation of teachers, exactly that which has caused so much difficulty in the area of evaluation, is no longer acceptable. As Hargreaves notes, change is a greater responsibility for teachers than the educational system: 'It is only by individuals taking action to alter their own environment that there is any chance for deep change' (Hargreaves 1997a, p.43; Stenhouse 1975, p.143).

Cohen and Barnes (1993), Darling-Hammond (1997), and Fullan & Hargreaves (1991) argue for a situation where teachers can work and learn new pedagogies in protected situations where

they can unlearn old practices and develop their skills with the new. ‘There is, after all, the possibility that their novice efforts would impede students’ learning, or would become unduly painful, or would damage them professionally’ (Cohen & Barnes 1993, p.247. See also: McLaughlin & Pfeiffer 1988, p.42; Scriven 1989, pp.98-99; Fullan & Hargreaves 1991, pp.72-73). In other words, evaluation ought to be a mechanism for supporting and defending the teacher in a time of vulnerability and to do this the mechanism must be able to monitor and foster the significant change taking place (McLaughlin & Pfeiffer 1988, p.5; Stenhouse 1975, p.169). Any current evaluative framework which does not foster such self-exposure necessary to implement change is already a threat to the pedagogical reforms being advocated (McLaughlin and Talbert 1993, pp.4-5; Darling-Hammond et al. 1986, p.240). Interestingly, the consultations in Queensland for the New Basics noted that ‘teachers view Education Queensland as a culture that discourages innovation and risk taking’ (Education Queensland 2000, p.22). For New Basics or any innovation to be truly effective the culture of the classroom needs to change, but so should evaluation procedures, to what McLaughlin and Pfeiffer call an *evaluation culture*, perhaps ahead of the change itself (McLaughlin & Pfeiffer, 1988, pp.15, 42).

A Matter of Professionalism

It is not sufficient to refer to what teachers do in the classroom as *professional* without them acting *professionally* towards the act of teaching. The report of the Australian Senate inquiry into the status of teachers defined a profession as made up of:

- a strong motivation or calling
- the possession of a specialised body of knowledge and skills acquired during a long period of education and training
- control over standards, admission, career paths and disciplinary issues
- autonomy in organising and carrying out their work
- the need for ongoing exercise of professional judgement
- members accept and apply a professional code of practice (Senate Employment, Education & Training References Committee 1998, p.23).

The first point would generally be accepted, but the vocational aspect has retarded societal respect and recognition and, perhaps, is the source of the relatively low remuneration compared to other professionals. It is readily cited in research on teacher attitudes (Scott et al. 2001). There may be exceptions, but few would see the long holidays as sufficient justification for

assuming the teacher's burden. The knowledge and training point applies in Australia where almost all teachers require specific tertiary training of between four and six years. In the United States, controversy has arisen due to unlicensed teachers being employed to fill shortfalls, with some arguing that the unlicensed teachers are as good as their licensed counterparts (Archer, 2002; Darling-Hammond 1997, p.21). The difficulty with the specialised knowledge criteria is that everyone has been to school and thus feels that they understand education and schooling, and are qualified to express a point of view. The autonomy in organising and carrying out work and ongoing professional judgement criteria apply to every classroom situation. The curriculum may be set, the classroom management procedures agreed upon, but ultimately the teacher applies these in his or her own classroom each day without interference from senior teachers or management.

The major issues for this study are the control over standards and acceptance of a professional code of practice, neither of which exist (Senate Employment, Education & Training References Committee 1998, p.26; Darling-Hammond 1988, p.68). There is no charter that teachers attest to on assumption of employment and the only entry standards in Australia are those set by universities for graduation from teaching courses, a fact that the poet A.D. Hope lamented as far back as 1967 (ACER, 2000, p.5). A pass in the relevant tertiary degree and/or diploma and teacher registration remain the only requirements:

In the case of teaching, new teachers are faced with a double jeopardy: the least amount of professional experience as preparation prior to employment and the greatest expectation of their level of ability to perform in the profession at entry. On appointment, they are immediately responsible for a whole class, full time. No other profession functions in this way (Ramsey 2000, p.107).

Teachers might agree to subscribe to particular philosophies of a private school or meet departmental guidelines for employment of civil servants, but there is no teacher profession based code (Ramsey 2000, p.31). Consequently, it can be argued that there is no ultimate set of principles against which teachers can be measured in any professional or ethical sense (Darling-Hammond 1997, p.213). Teacher registration procedures have been neither initiated nor assumed by teachers themselves. The externally mandated and conducted registration process merely verifies qualifications and checks criminal or administrative misdemeanours, not teaching expertise, which might influence employment of an individual. Ramsey noted:

In teaching, matters of professional conduct are the concern of employers rather than the profession acting through agreed standards of professionalism. Employers have developed codes of conduct and practice to provide a framework for professionalism in teaching. Teaching is unique when compared with the other professions studied, having no professional registration authority, no mandated system of continuing professional development and no professional oversight of teaching standards or practice (Ramsey 2000, p.99).

Put simply, teachers have no control over admission, standards, or practices within their own profession, because the teachers themselves have no mechanism by which to define or articulate them (Ingvarson & Chadbourne 1994, p.23; Kleinhenz & Ingvarson 2004, p.46; Ramsey 2000, p.117). Shanker notes a more rigorous process, at least in outline, in some areas of the United States with multi-tiered licensing systems and renewal of licensing dependant on achievement of development credits (Shanker, 1996, p.222). However, once again, these additional entry requirements do not appear to be set by teachers but by education boards or state regulation, and there remains debate on how effectively they are being applied (Kleinhenz et al. 2002, p.5).

Ingvarson and Chadbourne's extensive work on teaching standards for Australia has been based, initially, on filling this requirement. They argue that, without an agreed set of standards for what teachers are expected to know and do, regardless of their school system or curriculum, there is not a basis for making any judgement on the quality of the individual teacher's work or what they should know (Kleinhenz et al. 2002, p.6). It can be argued that, if teachers' work is isolated within the classroom and operating in an area beyond purely technical and bureaucratic frameworks, the ones to articulate the standards for teaching should be teachers themselves (Stenhouse 1975, p.143; Kleinhenz et al. 2002, p.24).

Whilst this is a strong argument, 'Cultural norms of privacy and individualism remain strong' (Kleinhenz et al. 2002, p.3. See also: Webb 1985, pp.83-4; Shanker 1996, p.223). Consequently, for any standards to develop, and, thereafter evaluations of their work to occur, teachers need to overcome their isolation and be willing to exchange something of their real teaching practices with their peers within a professional school community (Ingvarson & Chadbourne 1994, p.21; Fullan & Hargreaves 1991, pp.10-11). This would enable them to capture collectively the social-technical dimension referred to by Pusey (1976). In doing so,

not only would they be breaking down the barriers, but also developing a better understanding of alternative strategies and approaches to which they are not exposed currently (Gitlin & Smyth 1989, p.79). Such an approach would encourage teachers to participate in the debate about evaluation as well as enhance their receptivity to standards and evaluation processes of which they, the teachers, need to ultimately take ownership (Shanker 1996, p.224; Schon 1983, pp.328-329; Peterson & Chenoweth 1992, p.186).

By focussing the debate into the social-technological area as described by Pusey (1976), evaluation moves away from being something *done to* a teacher (Kleinhenz et al. 2002, p.8; Gitlin & Smyth 1989, pp.5-6; Peterson & Chenoweth 1992, p.1830). If teachers are to make a transition to something *they do*, the first step must be for them to be increasingly critical of their own pedagogy and willing to experiment to try and make it better, something more akin to the characteristics of a profession described above (Stenhouse 1975, pp.156-157; Elliot 1998, p.17). Schon's arguments for reflective practice and those encouraging such reflection and action research in the classroom can be significant contributors to this process. For teachers to be truly reflective, they must become client and outcome focussed and willing to adjust their pedagogy to achieve understanding and not just transmit content. If Schon's reflection-in-action is applied, when learning does not occur there is no question of allocating blame. Rather, the learning problem assumes central stage and the main objective becomes bringing the abilities of both student and teacher to bear on its resolution (Schon 1983, p.66; Darling-Hammond 1997, p.3). This may require the teacher to diverge from texts, bypass faculty mandated standardised tests, and seek new ways to impart understanding of the subject. Teachers who adopt such a reflective approach thus are not focussed on their own ego and protection of their entrenched practices. They are problem centred, and external assistance, which might involve examination of their practice, will be less threatening because such assistance offers a diversity of approaches to help the child learn (Peterson & Chenoweth 1992, p.182; Stenhouse 1975, pp.143-144). For a truly reflective teacher, student achievement and understanding is the reward.

Having focussed on the problem of learning, the interaction with colleagues with similar concerns or interests can become empowering and open up possibilities for faster resolution, or at least application of wider skills. Being thus focussed, teachers in action research groups - such as that described by Viadero (2002) and Distad et al. (2000) - appear excited and open, and become willing to reveal their classrooms in a process of diagnosis and experimentation. A teacher with a reflective perspective or engaged in action research is open to evaluation

because it represents not a threat, but a chance to bring to bear another objective and professional perspective on their pedagogy (Schon 1983, p.323; Stiggins & Duke, 1988, p.81; Darling-Hammond 1997, p.34).

It is not enough for the standards movement to establish its certification levels as in the United States. It must go hand in hand with the establishment of reflective practice to make the certification process a live developmental experience rather than an isolated academic exercise (Stenhouse 1975, p.162). Ramsey agreed in noting that:

written standards by themselves will not enhance teacher quality. The critical elements in increasing teacher quality and professionalism are the systems that will recognise, utilise and value these standards of professional teaching practice. Teachers will only value standards when the efforts they put into meeting them are recognised, valued and rewarded (Ramsey, 2000, p.146).

There is certainly evidence that the National Board for Professional Teaching Standards is encouraging teachers to form cooperative clusters to help them meet its stringent demands. However, unless the reflective process occurs concurrently, there is still that danger of the standards being *external* to the teachers.

A key aspect of the standards movement is the certification of, by, and for teachers. Ingvarson and Chadbourne argue that, for this to occur, there must be teachers who are both willing and committed to be part of the process, not just individually or through respective subject specific organisations, but as professionals (Ingvarson & Chadbourne 1994, pp.23, 27; Preiss 1994, p.270; Peterson & Chenoweth 1992, p.177). The fact that standards occur only in the Science and Mathematics subject areas, and these with little current application in Tasmania at least, suggests that much of the effort in evaluation remains within the tertiary and research field and does not appear to have been taken up actively by teachers themselves. Success requires both leaders and “foot soldiers” within the teaching community (Ingvarson & Chadbourne 1994, p.39; Ramsey 2000, p.142). Teachers with an independent reflective perspective will be in the vanguard.

Whilst establishing the connection between reflective teachers and the standards movement seems a logical process, and indeed desirable in making teacher evaluation a viable issue in the development of enhanced teacher professionalism, many obstacles remain. The most basic of these are the fostering of reflective practice in teachers, and countering the strong pressures for

retention of content focus in teaching. In fact, there is something of a “chicken and egg” issue in that we need more reflective teachers to provide the foundations for evaluation and enhanced teaching, and yet we need to take action to create the basis for effective reflection. This has not to do just with the structures in place, but also the perception of teachers themselves. Whilst the focus remains on content delivery, teachers can only be described as technicians. The syllabus lists what is to be taught, the students get their textbooks and are assigned the exercises to complete, and the teacher maintains order whilst the process occurs. External tests verify that the production process has been completed satisfactorily and that the requisite amounts of knowledge have been imparted to the student. At its worst, this represents the technical model of teaching, upon which ill-informed teacher competence measures in the United States are based; an inflexible bureaucrat driven, closed teaching process in which teachers only pursue volume and test success (Gitlin & Smyth 1989, pp.87-88; Tickle 2000, p.20).

The situation may not always be so clear-cut or negative. Teachers who follow formulae in their teaching may continue to generate students who achieve grades by applying knowledge and experience and time tested techniques (McLaughlin & Pfeiffer 1988, p.64). If nothing is done, the same teachers may well continue along the same track. Such teachers operate within an established paradigm built up, as Ramsey pointed out, on certainties of permanency and progression; a paradigm that meets the organisational solution to industrial revolution mass schooling (Ramsey 2000, pp.19-20). However, the issue becomes whether they have developed or questioned their pedagogy and the understanding of their students. Such a teacher may well argue that they are professional and apply professional judgement, but still act in a largely technical manner and do so successfully (Schon 1983, p.69). The school system is also likely to be supportive of such an established teacher. Indeed teachers may feel both appreciated and worthwhile within their schools, secure in their position as bureaucratic elements of the education process. Woods (1979) makes a useful distinction between such teacher-bureaucrats and teacher-professionals:

The former are more bound by institutional forms and processes and more geared to the formal definition of the teacher role. They are more likely to show a higher degree of rule consciousness, exert their authority, and foster formal and depersonalized relationships. They are categorized by pupils...as “too strict”, “full of moans”, “won’t laugh”, “treat you like kids”. Teacher-persons capitalize on humour and togetherness. They are “more natural”, “more like a friend than a teacher”, “have a laugh with you”. “talk to you

like real people”. They are still in control of the institution, using it for their ends. Teacher-bureaucrats, however, are governed by it (Woods 1979, p.244).

To create the expectation that teacher-bureaucrats will somehow change successful behaviours to become reflective or apply themselves to a demanding standards certification process seems naive as it presupposes the desire for change without some stimulus. It is the change in self-perception from technician to independent professional that must be the first step. Teachers may need to be encouraged away from their own self-perception as technician deliverers of the curriculum towards holistic educators of students, away from content to a true client-centred process of learning (Schon 1983, p.299). Without such a change there is unlikely to be incentive or motivation to seek standards certification and progress to a further evaluative process (Gitlin & Smyth 1989, p.163). The effort required may be shrugged off as excessive or “outside my duty statement”. As Ramsey notes, a major change is needed:

In too many schools in New South Wales, teachers are rewarded for waiting: for some by teaching on a “casual” basis until a permanent appointment is offered; through salary increments over the first decade of teaching; through accumulated transfer points in the government school system with the reward of eventual appointment to a school in a preferred location; or through the hold which a superannuation scheme may have in the latter years of teaching. This is a culture which rewards patience, not learning; it rewards maintaining the status quo rather than rewarding development of skills as a teacher. In a society where most areas of employment increasingly reward performance, creativity, innovation and adding value, the opportunities which teachers have to be recognised and valued appear so limited that they reflect a period which has long passed in much of the wider workforce (Ramsey, 2000, p.85).

If Kleinhenz, Ingvarson and Chadbourne (2002) are correct in describing Australia as a “blank slate” for evaluation, then it is highly desirable that teachers be encouraged to *do*, rather than just look at *how*. The standards movement has been criticised as being rather clinical and outside of the daily school environment where events and interactions of each day occur. As such its intellectual justification is unlikely to penetrate the daily life of teachers who will see its requirements as an additional burden. As mentioned above, teachers will need to develop a perception of themselves as constantly developing professionals and not process workers

(Schon 1983, p.231; Kauchak et al. 1985, p.37). This is a key challenge that will need to be addressed if the evaluation agenda is to move beyond academic research into reality and not face concurrent counter pressures as in the United States. In colloquial terms, teachers can be “shown the ball”, but they need to be motivated to pick it up, and to do that they need to see positive objectives. Once they have made the transition, the rewards should be sufficient for the process to develop its own dynamic (Peterson & Chenoweth 1992, p.182). Furthermore, in organisational terms, teachers need to be willing to take ownership of the social-technical dimension described by Pusey (1976) and enhance it to the level of maximising student outcomes as the foundation of schooling, not bureaucracy (Darling-Hammond 1988, p.73; Ingvarson & Chadbourne 1994, p.24). This is essentially the argument put by Darling-Hammond in her argument for changing schooling; once the focus has shifted, the organisation of schooling should follow the pedagogy and not the other way round.

Design Elements for Effective Evaluation

From the analysis in this chapter it is possible to deduce some design elements for effective evaluation. They are not mandatory or necessarily related to one specific form. Rather, they are framed from arguments and analyses of researchers, and based on those issues that have hindered evaluation becoming an integral part of the education process.

Purposes

(1) *There should be a clear distinction between the purposes of evaluation.*

Whilst this has been stated in the literature, it does not appear to be the case in practice. The conduct of “staff appraisal” by the school hierarchy, which also decides on promotions and tenure, actually works against the development of teacher practitioners by discouraging admission of anything that may be perceived as a weakness. If a formative evaluation is conducted, individuals must feel free to discuss and seek support for their most significant difficulties. Formative evaluation must be growth directed and objectively informative to the subject (Duke & Stiggins 1986, p.15). Similarly, accountability appraisals should be separated from professional evaluations aiming at the enhancement of teaching practice (Ramsey 2000, p.120).

(2) *Evaluation should promote and reinforce the desire of teachers to improve their practices.*

Teachers should be encouraged to embrace evaluation because it helps identify areas where they can improve. Evaluation should thus concentrate not on negatives or institutional

behaviours, but on improving professional knowledge and skills (Ingvarson & Chadbourne, 1994 p.13; Armiger 1981, p.298).

(3) *The objectives of evaluation processes should be positive and learner centred and expressed in such terms.*

The school is there for the students' learning, not the other way around. Evaluation should be framed from the same perspective, with judgements based on the positive contribution to student learning rather than hierarchical conformity. The pursuit of knowledge needs to be a joint activity rather than an imposed (Elmore 2000, p.20). If this is not the case and evaluation is perceived as a burden or punitive then teachers will withdraw from it (McLaughlin & Pfeifer 1988, p.62; Holly 1989, p.111; Gitlin and Smyth 1989, pp.40-41).

(4) *Evaluation should encourage teachers to be reflective and open to change not force them into greater isolation.*

Researchers see benefits in individuality for teachers, but isolation cuts teachers off from the benefits of interaction with peers. Teachers should be encouraged and rewarded for participating in evaluation and improving themselves. This means that they should be allowed and encouraged to take risks to move beyond habitual behaviours. (Fullan 1997, p.39; Schon 1983, p.335; Holly 1989, p.116; McNeil 1981, p.280; Gitlin and Smyth 1989, pp.163, 165).

Enabling Conditions

(5) *Teachers need to be active participants in the development of evaluation.*

Organisational theory suggests that any imposed system will not only generate anxiety, but also fail to target effectively the necessary areas for consideration. With the ultimate aim being improved teaching and learning, teachers should engage and support the system in their own interests as well as those of the organisation and the students (Kleinhenz et al. 2001). McLaughlin and Pfeifer describe this as the establishment of an *evaluation culture* (McLaughlin and Pfeifer 1988, pp.15, 26-27, 81; Stiggins and Duke, 1988, pp.22-23; Armiger 1981, p.299).

(6) *Teacher evaluation should not be conducted in isolation but be part of a school-wide quality assurance program.*

It is the duty of teachers to teach and for administrators to provide them with the means by which to teach effectively (Stiggins and Duke 1988, pp.16-17; Stake 1989, p.17). It makes no sense for teachers to be evaluated without making judgements about the resources or

management of the school. If money is misspent, or students force to learn in overcrowded conditions with inadequate facilities, or if teachers are prevented constantly from accessing up to date materials, then it is unfair to judge their teaching as if these were not limitations, particularly if the evaluation is summative. Similarly, teacher support for school goals cannot be judged if such goals are unclear or interpreted differently within various parts of the organisation. Furthermore, demonstrating leadership accountability highlights that evaluation is considered as a serious part of the school culture (McLaughlin and Pfeifer 1988, p.44; Stiggins and Duke 1988, p.17; Stake 1989, p.13; Scriven 1981, p.248; Holly 1989, p.116; France 1989, p.77). From a school governance perspective, such accountability is also an essential element of leadership (Elmore 2000, p.32).

(7) *Summative evaluation is inappropriate in the absence of any formative evaluation process.*

It is manifestly unfair for a teacher to be judged as lacking in any way without being given the opportunity to develop capabilities or skills necessary for that particular school environment. Weaknesses, shortfalls, and inadequacies should be identified and the opportunity for rectification provided. Those seeking promotion or tenure should know what they are doing right or wrong and be given the opportunity to either reinforce or rectify such behaviours or skills. Scriven argues that no formative evaluation can exist without it having a summative level because of the ultimate accountability requirement (Scriven 1994, p.73). The relationship between the two should be established and clearly stated.

(8) *Evaluation should not be an isolated event but occur frequently enough to allow early rectification of problems or encouragement for work well done.*

Organisational theory tells us that frequent feedback is preferable to widely spaced evaluations that may refer to events isolated by time. This allows reinforcement of positive behaviours, increased motivation, rectification of difficulties, and monitoring of remedial action (Stiggins and Duke 1988, p.17; Holly 1989, p.112). One evaluation a year cannot help but be perceived as a summative annual report. More infrequent events are unlikely to be treated seriously unless they involve promotion or salary increases.

Structure

(9) *Evaluation must examine the work and not the person.*

Pusey's (1976) organisational analysis helps us to recognise the difficulties of evaluating teachers because of their transference into the social domain and a readiness to perceive

criticism of self. The style of teaching, even if it is completely antagonistic to that espoused by the evaluator should be valid provided that it is ethical, is accepted by students, and achieves positive learning outcomes (Scriven 1994, pp.82-83; Bednall 1989, p.50).

(10) *Evaluation must be based on trust, mutual confidence, and professional equality.*

Teachers must be able to trust those who conduct summative evaluation to have the relevant expertise, to be open, fair and objective and to believe that the outcome will be of value (McLaughlin & Pfeifer 1988, pp.4, 15, 49, 66; Stiggins and Duke 1988, pp.22, 85; Duke & Stiggins 1986, p.25; Armiger 1981, p.298). With formative evaluation, they must also feel secure enough to admit shortfalls and seek assistance to remedy them. As soon as trust is lost, because of inappropriate procedures or misuse of information, the evaluation environment is compromised. It follows that the same evaluator should not judge a teacher for both formative and summative purposes.

(11) *Evaluation should be site specific and meet the needs of teacher and the students within the school community.*

There is no generic set of competencies. The needs of students will vary between schools in different environments under divergent education systems. Beyond just content knowledge, teachers will require different emphases in their skill repertoire to generate learning. What works in one school not only may be not transferable but completely inappropriate. This does not require a “reinvention of the wheel” on every occasion, but some modification should be made to take account of local requirements (Holly 1989, p.112).

(12) *Evaluation of teachers should be discipline specific.*

With content knowledge a key requirement and the direction of specific pedagogies to support learning, teachers of a subject, or a specific teaching area such as early childhood education, should be evaluated with respect to that discipline (Stiggins and Duke 1988, pp.83-84; Holly 1989, p.112; Armiger 1981, p.299; Bednall 1989, p.48). Teachers forced to teach in areas beyond their expertise cannot be evaluated fairly if they lack either content knowledge or specific pedagogical techniques accepted as necessary for the task.

(13) *Duties should be included in addition to academic evaluation of pedagogical skills.*

Scriven’s Duties of The Teacher (DOTT) is a good starting point to expand beyond the classroom to the broader responsibilities of the teacher. Capabilities in and beyond the schoolyard are part of the broader teaching role and should be given appropriate weighting in

any evaluation (Scriven 1981, p.250; Darling-Hammond et al. 1986, p.227). As Scriven points out, the issue is one of teacher's worth in a school, rather than just merit that can be accumulated as qualifications (Scriven 1981, pp.246).

(14) The evaluation process should be dynamic and be able to respond to changes in the social and educational environment of schools and teachers' developing roles.

If change is a constant, then evaluation procedures cannot be continued without considering new duties, responsibilities, and approaches to education. Without such updating, formative evaluation in particular may miss assisting the teacher to develop the attributes required in the changed process, thus undermining the desired outcome (Hargreaves 1997a, p.9).