two paradigms of ASSESSMENT

Paul Hager (University of Technology, Sydney)
&
Jim Butler (University of Queensland)

HAGEP94.234

Abstract

Many educational developments in recent decades pose a serious challenge to the traditional scientific measurement paradigm that has dominated assessment practices. The scientific measurement paradigm has led to an over-emphasis on statistical tests and the reification of single measure test scores. The educational developments that challenge the scientific measurement paradigm include problem based learning, newer understandings of cognition, and the rise of performance assessment. These developments reflect widespread attempts by educators to reform assessment practices so as to encourage more effective learning. As a result, a new paradigm of educational assessment, which we call the judgmental paradigm, is emerging. The basic assumptions and features of these two paradigms will be compared and contrasted by referring them to a three level conceptual model of education, training and assessment for workplace performance. In this model, each of the levels is nested in, and a prerequisite of, the next level. Also each of the levels has its own appropriate curriculum and assessment technology. The contrasting strengths, limitations and appropriate uses of the two assessment paradigms are underlined and clarified by considering their differing contributions to, and roles in, the proposed three level model of education, training and assessment for workplace performance.

Although educational assessment has been dominated by a paradigm of scientific measurement, a wide range of educational innovations and initiatives over the last twenty years have seriously challenged the value and rationale of these traditional assessment practices. These recent innovations and initiatives include problem based learning, education for capability, portfolio-based performance assessment of teaching (Shulman 1987), the growing focus on the assessment of clinical competence in the final stages of medical and paramedical courses, and the increasing concern by cognitive theorists to assess how students go about solving problems as well as the solutions that they reach. We believe that these developments reflect the rise of a new paradigm of assessment. Traditional assessment practices are based on a paradigm whose features are well known and understood, which we call the scientific measurement paradigm. The new paradigm, which we call the judgment paradigm, is not so well known or understood in educational circles.1 However the making of a judgment on the basis of

a collection of evidence is familiar in both legal processes and in the workplace. The nature of this new assessment paradigm is the main focus of this paper. However it will be helpful to start by examining some wider factors that have contributed to the change in our thinking about educational assessment and thereby created the climate that has encouraged the emergence of a new paradigm of assessment.

Factors Leading to a New Assessment Paradigm

We are involved in a global shift in understanding of the nature of working. This statement applies to adults working at all levels and in every occupation. Conceptions of the mechanisms that underlie our thinking and reasoning processes have altered dramatically during the 1970s and 1980s. Our thought processes are not what we thought they were. This reconceptualisation of our mental lives, of our ways of knowing the world, of our ways of developing excellence in our lives and our work, has led to new understandings of how to enrich them through education and training. (See Butler 1992, 1994). Recent developments in cognitive science which are part of the reconceptualisation, are providing new avenues for designing the goals and processes of education and training.

In essence the shift comes down to two approaches to understanding people's capacities founded, respectively, on the concepts of intelligence and cognition. Intelligence is the measure that has been widely used to determine the level of learning and working that people are capable of achieving throughout their lives. According to the intelligence approach, the learner is conceived of as an individual with a genetically programmed limit. This concept leads to the categorisation of people which underpins a hierarchy of power and influence. This categorisation of people thereby supports highly segmented and specialised roles within the workplace. It has been used in the workplace to predict the learning expected of people and to understand the types of work that they are fitted to do. The process of human development within the intelligence approach is one which takes place within the bounds of the individual person, with limited influence by factors outside the person. The basic assumptions surrounding the intelligence concept have been strongly contested by empirical studies and by the theories of cognition that have become dominant over the last twenty years (Hayes, 1990). Although suspicion about the significance of IQ scores is now much greater than it once was, the continuing influence of the intelligence concept is evident in the widespread faith that is shown in Tertiary Entrance Ranking scores (or their equivalent) as predictors of ability to do almost anything. The cognition approach, on the other hand, focuses on the process by which we think and reason. It leads to the categorisation of tasks, of the thinking and reasoning processes and of the products of thinking, but not to the categorisation of people. Cognitive models posit

considerable scope for individual capacity to develop and grow through the interactive processes between persons and varying contexts (Hayes 1990; Yates and Chandler 1991).

Impact on the Preparation of Professionals

In many countries, the assessment of professional practice, skill levels or occupational competencies is an area that is in a state of intense development (Maatsch et al, 1987; Hager et al, 1994). Despite the fact that it would be extremely rare for people to answer a battery of multiple choice questions in everyday practice at work, and even more unusual for them to practice a specific skill in isolation, assessment technologies often rely on these techniques (Lareau, 1986). The American teaching profession is one such example. The national testing schemes for the registration of teachers, based on multiple choice formats, have undergone intensive development and have been extensively studied. A recent group of reviewers (Haney et al, 1987) conclude with this forthright statement:

no evidence has been provided to indicate that currently used tests are significantly related to general teacher competence or performance. Therefore we conclude that people who refer to currently used teacher tests as "teacher competency tests" are either engaging in self delusion, in sloppy use of words, or in outright fraud.

This is but one illustration of the widespread dissatisfaction with the professional assessment procedures that have been developed and used

for many years (McGaghie, 1989; Dawkins, 1989; Masters & McCurry, 1990; Gonczi et al, 1990). There are at least four reasons for this dissatisfaction. Firstly, the traditional assessment procedures typically sample a very narrow range of contexts compared to those which the practising professional is likely to encounter. The need to ensure a relationship between the professional practice context and assessment technologies for registration is well documented (McGaghie, 1989; Salman, 1981). Secondly, the assessment procedures are biased towards the assessment of practitioner knowledge rather than the performance capabilities of the practitioner. Thirdly, where the skills of candidates are assessed, this is done by experts who usually have other responsibilities in the practice setting and who therefore use indirect, vicarious and unstructured methods. Lastly, within the scope of the dominant assessment paradigm virtually no attention is given to the assessment of the personal traits or attitudes of those candidates being assessed.

For these reasons, assessment technologies are attempting to move away from an emphasis on the assessment of knowledge and are searching for schemes more closely aligned to the complex practices and procedures that the person will be expected to carry out safely and skilfully. The

goal is to ensure that the criteria of success in education or training accreditation processes are the same as those used in the practice setting. This new paradigm has been termed the judgment paradigm (Hager & Butler, 1994; Hager et al, 1994).2 Most of the explorations for new assessment procedures are presently focused on competencies schemes (Boyatzis, 1982; Benner, 1984; Burke, 1989). The identification and assessment of competencies is now widespread within the professions, the paraprofessions and many other occupations (Gonczi et al, 1990).

A Model of Education, Training and Assessment for Professions and Other Occupations

In order to prepare entrants well for a profession or occupation, it is necessary to have a conceptual model that encompasses the best understanding of what is involved in the practice of the profession or occupation. This model is necessary to guide both the curriculum development and the assessment technology. The model offered in this paper has three levels with each level nested in and a requirement for the next level. Each of the three levels has an appropriate curriculum and assessment technology. The order of the levels, one to three, can be described on many separate dimensions:

- component parts to the whole
- analysis to synthesis
- aspects remote from practice to actual practice
- external rules governing behaviour to personal, experiential integration of practice
- achievements early in the curriculum to exit achievements.

Each of these dimensions provides guidance to designers of curriculum and assessment procedures.

The three levels of this model for education and training will now in turn be outlined and discussed:

1. Knowledge, Attitudes and Skills

This is the level that historically dominates the education of professionals, but is less prominent in the training of skilled artisans. The knowledge section of the curriculum is derived from the sciences and social sciences that are regarded as being foundational to

practice. This is the context-free knowledge that the curriculum designers decide is needed for future practice. There is also a skills section of the curriculum that is designed from a list of all those isolated skills - psychomotor and attitudinal - that the effective practitioner is deemed to need.

Thus, the curriculum becomes an orderly sequence of knowledge and

cognitive, technical and interpersonal skills. The assessment instruments are usually made up of:

- multiple choice tests to measure retention of relevant knowledge
- multiple choice tests to measure cognitive skills
- problem solving exercises to measure skills in this area
- observational checklists of skills in the practice setting.

There are those who argue that skilled practice is totally reducible to knowledge, attitudes and skills. They develop a "Knowledge and Skills" conceptualisation of education, training and assessment. However, the congruency between this conceptualisation of the education and assessment of practitioners and the real world of practice does not appear to be very great. It has been shown (Maatsch et al, 1987) that when highly trained and expert practitioners assess their peers in the practice setting they do not perceive the performance in the terms of the "Knowledge and Skills" model. Rather, expert practitioners grasp the general level of competence displayed in the total practice. This finding is important because as the researchers point out:

Expert examiner evaluations of physicians in their own speciality is, after all, the only "Gold Standard" that educators, researchers, the public, the courts and physician colleagues alike will accept.

Thus, this first level of the model of practice is accepted as necessary but rejected as sufficient. It is designated as the first, earliest and most remote level that needs to be supplemented by the following two levels.

2. Performance in Simulated or Practice Domains

Performance is a higher level of integration of knowledge and skills, which presupposes the attainment of sections of the previous level. The student is placed in simulated or practice settings to perform detailed procedures at a more macro level. The performance is still conducted in a contrived setting, but it does require the student to operate at a level that integrates several knowledge areas and skill regimes. The curriculum for this level is necessarily derived from knowledge areas and performance roles that are proximate to the practice area of the professional candidate.

The assessment of performance is widespread in the form of performance appraisal checklists. Typically, they are used with little documentation to assist their application and little reference to the context in which the performance occurs. These stark lists of performances are scored as "satisfactory" or "unsatisfactory". The lists tend to be impersonal and fragmented. There are so many procedures that it is difficult to ensure that all are covered.

Secondly, there is usually little guidance as to the process by which the global decision about a student's performance, based on numerous "satisfactories" and perhaps a few "unsatisfactories", will be made. And the relationship between this global judgment and level of practice skills is left unexamined.

The level two approach to the design of the curriculum for education, training and assessment for practitioners recognises the existence of a more integrated, macro level of performance. However, it still holds to an underlying assumption that practice can be captured in a list of isolated performances. This paper argues that such an assumption is unlikely to provide valid assessments for decisions concerning the skill levels of practitioners.

3. Personal Competence in the Practice Domain

The third level of the model - personal competence - is a characteristic that can only be displayed in the practice setting. Therefore it can be learnt and assessed only in the practice setting. It demonstrates the highest level of personal integration of knowledge, skills, personal qualities, and accumulated reflective experience. It includes the previous two levels but goes far beyond them in the sense that their contributions are at the service of higher order thinking in a setting of complex practical problems. It is this form of reflective and integrative thinking that needs to be the focus of assessment procedures that are faithful to real practice.

Personal competence is a construct whose definition is debated and confused (Klemp, 1980; Burg et al, 1982; Benner, 1984; Mansfield, 1989; Hager, 1994; Hager et al, 1994). The definition of competence, proposed in this paper, is the ability of a person to fulfil a role effectively. It is an ability that encompasses the entire range of demands that make up a complicated role. Competence includes being able to carry out a range of routine and predictable procedures, having a store of knowledge and being able to pass knowledge-rich examinations, but it is much more than all of that. A competent practitioner is also expected to be flexible and versatile, a reflective practitioner, a manager of change who is willing to innovate, and a person who has the attitudes and motivations to act skilfully and ethically. Inherent in this definition is the expectation that the competent practitioner effectively demonstrates the ability to practice over an extended period of time within a wide range of diverse contexts that include uncommon occurrences and contingencies. Concerning the curriculum at this level, Kermode (1987), in calling for learning in the practice setting, refers to the learning richness of this setting and the contextual subtleties that impinge on the process of practice.

Implications of the Three Level Model for Assessment

In occupational fields, the validity of the assessment technology used for measuring attainment of skill levels is essential for public and professional accountability and safety (Fahy & Lumby, 1988). Education and training for professions and other occupations is often split between two settings: the practice setting with its clinical skills tests (Fahy & Lumby, 1988), or informal observations; and the tertiary education institution tradition, with its reliance on examination rituals (Morgan, 1989; Ashenden, 1990) Establishing satisfactory links between assessments in the two settings is an ongoing concern.

Simulations have been suggested as a compromise between assessments which concentrate on knowledge and reasoning processes and assessments carried out in clinical settings, where uncontrolled variables may disadvantage candidates. McDowell et al (1984) found that performance in simulation assessments with real patients correlated well with assessment ratings in patient care settings. The correlation was lower with electronic patients. However, simulations, even with real

patients, refine and reduce the focus of assessment in order to ensure that the criteria are specific and thereby remove much of the ambiguity and uncertainty not to mention the other stimuli which impinge on performance in the clinical setting.

Practice readiness assessment - the interpretation of data derived from assessment of practice (McGaghie, 1989) - is another alternative which achieves the aim of relating assessment to practice. It is also based on the judgment paradigm discussed earlier. When the clinical setting is used for assessment, the constantly changing environment and the ambiguity and multiple stimuli of this setting pose authentic problems for the candidate to tackle and complex difficulties for the assessor.

The complexity of the interrelationships among candidate, client, environment and assessor makes the task of separating the intrinsic standard of performance of the candidate from these other factors very difficult. On the other hand, alternate means of assessing professional competence by attempting to reduce the complexities and thus simplify the assessment task, reduce the validity of the assessment. It is this tension between reliability and validity that needs to be carefully and openly balanced. There are published observational technologies applicable to simulation settings or real settings with real patients or patient actors with real or simulated diseases (Cason & Cason, 1984; Maatsch et al, 1987; Ross et al, 1988; Conn & Cody, 1989). The research reports on criterion-referenced rating scales in clinical settings are encouraging (Bondy, 1983; 1984).

The human assessor is the basic common element that gives coherence to the data from these various sources of evidence for assessments based

on the judgment paradigm (Masters & McCurry, 1990). The qualities of the human assessor are central to the reliability and validity of the whole process. Therefore, to complete the assessment technology, once all of the data gathering techniques have been decided, it is necessary to devise:

- standards of performance on each competency; and
- mechanisms for communicating these standards to assessors.

Standards for Competencies

In any assessment technology based on competencies, each competency requires a precise standard. There is little published research that helps solve the standards problem (Sadler, 1987). Wandelt and Slater Stewart (1975) based their scheme on the premise that all nurses held a common standard of the level of performance expected of a competent beginning nurse. However, they did not state the details of this standard. On the other hand, in a quite sophisticated assessment study, a very arbitrary standard was used; a failing score was defined as "two standard deviations below the mean score" (Conn and Cody, 1989). These types of standards are not applicable to competencies schemes.

One global manner by which standards can be stated is to describe, e.g., the beginning nurse level in terms of the student and expert levels of practice. This formulation of the standard is presented in Table 1.

TABLE 1: Specifying Minimum Standards: The Newly Registered Nurse

Student Nurse Newly Registered	Works under supervision in structured settings
Nurse Expert Nurse	Works independently, and interdependently in semi-structured settings
	Works independently, and interdependently in un-structured

settings

with predictable client outcomes
with predictable client outcomes
with unpredictable client outcomes

to non-complex problems.
to non-complex problems.
to complex, multiple problems.

Knowledge is accurate but partial.
Knowledge is comprehensive and accurate.
Knowledge is comprehensive, accurate and tacit.

Because of guidance, interventions are always safe, accurate and appropriate and always have safe outcomes.
Interventions are always safe, appropriate, accurate and always have safe outcomes.
Selects the best interventions which are always safe, and are performed safely and accurately.

The standards were derived by describing the expected level of practice of a student nurse, a newly registered nurse and an expert nurse. The level of practice for each is specified in two dimensions: the level of complexity of the nursing problems in the practice setting and the level of independence and expertise of the response to nursing problems.

Communicating Standards to Assessors

As well as stating the standard there must also be a means of obtaining shared understanding of that standard by a wide range of assessors. Tacit knowledge of standards held by expert assessors needs to be brought into consensus. This sharing allows assessors to agree on the standards on the individual competencies and the meaning of minimum beginning level competence for registration. The implementation of the competency-based readiness assessment process depends on the willingness of assessors to undertake training in the purposes, possibilities, procedures and pitfalls of the assessment technology.

Assistance given to assessors is firstly in the form of guidance as to the possible sources of evidence for each competency. Secondly, there is a list of possible "cues" for each competency. Cues are selected concrete examples of activities illustrative of the competency.

Communicating standards of performance to assessors and the sharing of standards by groups of assessors can be achieved by ensuring that the following matters are carefully attended to:

- The assessors are expert in the practice that they are observing.
- The assessors are trained in the techniques of observational assessment: self understanding as a component of judgment; where to gather evidence; the effects of context on assessment; observational skills; interviewing skills including the development of questionnaires; item construction skills; skills in interpreting evidence; the scope and level of beginning practice to be assessed in relation to their own.
- The standards applicable to each competency are given a precise verbal description that conveys as much meaning as possible, and is under constant review. The assessors continue to review the formulation.
- The assessors are given detailed cues and explicit instances of the competencies - critical actions checklist (Maatsch et al, 1987) - that can lead to the inference of the competency. The assessors also continue to add to and modify the lists of cues.
- The assessors are given extensive experience in assessment, so that they become expert in this important function.
- Groups of assessors engage in activities by which they can share their standards, discuss cases and converge on some consensus about the standard on each competency, so called "moderation meetings" (Fitzpatrick, 1989).

Research results from very large and intricate studies have shown that it is possible to obtain quite high inter-rater reliabilities with highly trained assessors (Maatsch, 1987; Conn & Cody, 1989).

We will conclude this section on the assessment of performance against standards with some general comments:

- All assessments technologies associated with knowledge, skills, performance, thinking skills values, competencies, etc, have drawbacks.

No assessment technology can examine the full range of knowledge or competence. The range of subject matter selected depends on the perspective of the examiner in the case of written or simulated tests, and on the exigencies of the clinical setting in the case of observation of competence. With written or oral testing it is possible to assess what the candidates say they will do but not what they will actually do. The information obtained from such tests is also dependent on the skills of the writers of the examination items and on the criteria for marking.

•Observations of competencies are expensive as they involve one-to-one

assessment. They also are dependent upon the skills of the observers. The success of the interview in eliciting valid information depends on the self knowledge and honesty of the interviewee and the skill of the interviewer. Moreover, the lack of shared standards and meanings with clients and peers makes evidence collected from these people difficult to interpret.

•In any of the forms of assessment discussed above, the information gathered must be interpreted in terms of competencies and standards. This raises the question of the subjectivity of the assessors. Objectivity tends to be equated with low inference numerical data, but this is of very limited use in the assessment of competencies. Objectivity is often contrasted with subjectivity portrayed as high inference judgments, but this contrast is facile (Deutscher, 1983). Subjectivity can never be eliminated from the assessment process. It is also undesirable to do so. Judgments about competence depend on the tacit knowledge and expertise of the assessors. This does not mean that subjectivity is equivalent to bias. It is the learned, relative standards of the assessor being used as the basis for assessment judgments. Moreover, if standards can be learned they can also be shared and brought into some relationship of uniformity with the subjective standards of others with equivalent knowledge and experience. Objectivity is the intelligent learned use of subjectivity, not a denial of it. It is the assessor who delivers objectivity, not the data.

•The reliability and validity of an assessment technology based on competencies is a key concern. It is under extensive analysis at the present time, and will need to be continually addressed. It must not be forgotten that the assessment methods applicable to knowledge and skills are just as problematical in their application.

•Competency-based readiness assessment is not a fully developed working system. As well as the difficulties noted above, there are the additional risks (not confined solely to this mode of assessment) of narrow interpretation of competence, and of the competencies and associated assessments becoming unrepresentative of current practice.

These limitations can, however, be avoided by providing for regular review of the list of competencies and of the assessment procedures by open democratic processes. Such reviews can also help to ensure that gender, ethnic and migrant issues are monitored.

Fundamental Assumptions and features of the Two Assessment Paradigms

Thus far in this paper, the traditional scientific measurement paradigm of assessment has been contrasted with the emerging paradigm which we call the judgment paradigm. Some applications of the newer paradigm to assessment in the professions has been discussed. Finally, we would like to provide a more detailed comparison of the two paradigms. We begin with what they assume about the nature of the human subjects that they are measuring. The basic metaphysical assumptions about humans that underpin these two different approaches are set out in Table 1 below.

Table 1
Metaphysical Assumptions about Humans

Intelligence Approach	Cognition Approach
Human capacities fixed by ‘g’ factor	Human capacities reflect complex combinations of knowledge, skills, abilities, etc. Some fixed, some not.
Single factors	Diverse factors
Normal distribution of ‘g’	Distribution of complex combinations of knowledge, skills, abilities, etc. that are not assumed to be normal
‘g’ factor underpins all performances	Different kinds of combinations of knowledge, skills, abilities, etc. underpin different kinds of performances

These two contrasting philosophical positions have been characterised by Hayes (1990: 236) as 'preformationism' vs. 'epigenesis'. In preformationism "...all the structures are assumed to be present from conception, waiting merely to unfold, mechanically, across the lifespan of the person...", while in epigenesis "....the organism is seen as containing only the basic elements" with "....considerable scope for modification by the processes of interaction between person and environment." Thus the two contrasting positions have distinctly different implications for the role of education and training and the types of assessment that are employed therein. The cognitive view

argues that human development is not the gradual drawing out of our individual potential as measured by our IQ, but, rather, from the earliest stages of development there is cognitive complexity to life's tasks and all associated learning. Cognitive theories lead to quite different possibilities for the relations between adults and work. Whereas the intelligence approach encourages selection of people to fit prespecified jobs, the cognitive approach enables us to view the workplace as a set of opportunities for people to learn and grow. It is this shift in understanding from the metaphysics of intelligence to the metaphysics of cognition that we are witnessing today. It is the shift in emphasis from the limits of thinking, working and learning to the processes of thinking, working and learning. This shift is one of the explanations for the increasing interest in problem-based learning, education for capability, portfolio-based performance assessment, the work-competency movement, etc. These educational innovations and initiatives are being more widely adopted and, taken together, they constitute a formidable challenge to the dominant scientific measurement paradigm of assessment.

Basic features of the two assessment paradigms

The hallmark of the scientific measurement paradigm has been a focus on measures that are maximally objective. This quest for scientific objectivity in testing has led to an over-emphasis on statistical tests and the reification of single measure test scores, with IQ testing providing a case study of the abuses that can result from such thinking. When assessment is driven by the imperative of scientific objectivity in testing results, assessment methods which best yield such objectivity, e.g. pen-and-paper tests conducted in controlled conditions, are strongly favoured. The results of this have included:

- theory being seen as easier to assess than practice (and students being given the message that theory is thereby more important than practice)
- assessment of well-founded certain knowledge
- knowledge being seen as objective and context independent
- assessment based on closed problems with definite answers avoiding more open problems with indefinite answers
- the traditional disciplines being viewed as the major part of the curriculum

What characterises problem-based learning, and the other initiatives and innovations listed earlier, is their adoption of a judgmental paradigm of assessment to replace the scientific measurement one. These are all instances of widespread attempts by educators to reform assessment practices so as to encourage more effective learning (Wolf et al. 1991; Ewell 1991). The adoption of a judgmental paradigm of assessment leads to quite different results from those produced by the scientific measurement paradigm:

- assessment centres on the integration of theory and practice
- assessment recognises the importance of provisional knowledge to practice
- the stress on assessment of practice draws attention to the subjective and context- dependent aspects of knowledge
- assessment, like actual practice, includes open problems with indefinite answers
- the curriculum is conceptualised as interdisciplinary and problem focused

A basic outline of the different assumptions about the nature of knowledge that underpin the two assessment paradigms is given in Table 2.

Table 2
Epistemological Assumptions

Scientific Measurement ParadigmJudgmental Paradigm

Theory prior to practiceIntegration of theory and practice

Foundationalist knowledgeNon-foundationalist knowledge

Certain knowledgeProvisional knowledge

Impersonal knowledgeKnowledge as a human construct

Context-free knowledgeContext-dependent knowledge

Disciplinary and subject focusInterdisciplinary and problem focus

Structured problemsUnstructured problems

Curriculum unproblematicCurriculum problematic

The nature of the measurement process according to the two paradigms

The scientific measurement paradigm centres on the production of numerical scores that constitute evidence about examinees' innate capacities. Hence, the measurement process is designed to ensure valid and reliable measures. In practice the main emphasis has been on reliability, i.e. on score production that is scientifically reproducible. The result has been scientific technologies of measurement which have tended to become ends in themselves in education. One indication of this is the widespread lack of understanding amongst educators that a different assessment paradigm is possible. Yet, as Norman, Van Der Vleuten and De Graaf (1991) amongst others have found, assessment techniques that focus mainly on producing

scores that are scientifically reproducible too often trivialise the curriculum content that they are supposed to be assessing.

As against the scientific measurement paradigm that has dominated traditional testing, the judgmental paradigm of assessment underpinning assessment of the various educational initiatives and innovations, listed earlier, is a quite different one, more akin to a legal paradigm (Bailey 1993: 6). For example, the decisions made in the judgmental paradigm about a candidate's performance are more like those that are made in a court of law, than they are like decisions made on the quality of output from traditional examinations. Thus a judgmental paradigm of assessment allows for the calling of more evidence in a doubtful case, rather than relying on making inferences from a fixed and predetermined set of data. Another possibility, allowed for by the judgmental paradigm, is that assessment can become a dialogue between the person being assessed and the assessor, so that there is scope for the person being assessed to present their case. In this way, learning and assessing can become much more integrated than they have been under the scientific measurement paradigm. The judgmental paradigm also creates interesting possibilities for the use of some self assessment. The integration of learning and assessing also has important implications for the current educational ideal of lifelong learning.

The different views of the nature of the measurement process adopted by proponents of the two paradigms are shown in Table 3.

Table 3
Assumptions ABOUT the NATURE OF THE measurement process

Scientific Measurement ParadigmJudgmental Paradigm

Measure ‘g’ factorInfer kinds of combinations of knowledge, skills, abilities, etc. from performances

Problem of objectivity:Problem of induction:
i.e. how to ensure data is scientifically i.e. how to collect sufficient
reproducibleevidence

The different assessment principles and practices recommended by the two paradigms

Eisner has noted the increasing interest amongst educators in "more authentic approaches to educational assessment" (1993: 219). He goes on to outline and discuss eight principles or criteria that distinguish the emerging approaches to authentic assessment from more traditional

approaches to assessment.3 In fact the Eisner principles serve very well to distinguish what we have called the judgmental paradigm of assessment from the scientific measurement paradigm. The extent of differences in the principles that characterise the two paradigm seems to us to be great enough to justify our claim that a paradigm shift is occurring in educational assessment. Once again the principles that characterise the judgmental paradigm are clearly very close to those that underpin assessment in the various educational innovations and initiatives listed earlier in the paper. These contrasting sets of principles, adapted from Eisner, are shown in Table 4.

Table 4
Assessment Principles

Scientific Measurement Paradigm	Judgmental Paradigm
Assess tasks remote from the world outside of classrooms	Assess tasks that reflect outside contexts
Assess solutions to problems	Assess processes by which problems are solved, as well as the solution
Assess simplified, discrete tasks	Assess performance of holistic tasks in their actual context
Assess individuals only	Assess group work as well as individual work
Emphasise one right solution	Emphasise alternative ways to reach acceptable solutions
Assess tasks directly from the curriculum as taught	Assess tasks that are relevant to the curriculum but expand on it
Assess discrete tasks one by one	Assess performance on holistic tasks as well
Teachers rigidly prescribe nature and form of assessment tasks	Learners help to design nature and form of their assessment tasks

Finally, it may be useful to emphasise major differences in the assessment practices that the two paradigms would employ. These are summarised below in Table 5. Most of the practices listed in Table 5 derive from the foregoing discussion, though the desirability of evidence from a variety of assessment events in the judgmental paradigm perhaps needs more comment. Just as legal judgments usually require a range of evidence from a variety of sources to make the inference safe, likewise for educational assessment based on the judgmental paradigm.

An example might be the judgment as to whether a candidate at the completion of a problem-based medical course is competent at making diagnoses. The decision properly would be based on evidence of various kinds from various sources.4

Table 5
Assessment Practices

Scientific Measurement Paradigm	Judgmental Paradigm
Examination focus	A variety of diverse assessment events as the focus
Measurement of attainments	Inference of competence
Single scores and measures	Multiple sources of evidence
Controlled test conditions	Simulation of life situations
Emphasis on objectivity	Emphasis on avoidance of bias
Focus on test instruments, validity, reliability	Focus on triangulation, direct evidence, informed judgment

Conclusion

This paper has argued that there are two distinctive paradigms available for educational assessment and that only one of them, the judgmental paradigm, is suited to recent and, increasingly, widespread educational innovations and initiatives such as problem based learning, education for capability, and portfolio-based performance assessment. This is evident from the contrasting features of the two paradigms that have been outlined in the last part of this paper. As the influence of the newer judgmental paradigm increases, a more favourable climate for the wider adoption of these and other similar educational innovations and initiatives will be created. This process will likely be accelerated by proponents of these educational innovations and initiatives extending their mutual professional contacts.

REFERENCES

Ashenden, D. (1990) Employment and Skills Formation Council: The Recognition of Vocational Training and Learning. Commissioned Report No. 1. Canberra: National Board of Employment, Education and Training.

Australasian Nurse Registering Authorities Conference (1990) ANRAC National Competencies for the Registration and Enrolment of Nurses in

Australia. Adelaide: Nurses Board of South Australia.

Bailey, M. (1993) 'Theoretical Models and the Assessment of Competency'. Paper presented at the VEETAC National Assessment Research Forum, Sydney, April.

Benner, P. (1984) From Novice to Expert. Menlo Park, CA: Addison-Wesley Publ. Co.

Boggs, D., Baker, B. & Price, G. (1987) 'Determining Two Levels of Nursing Competency', Nursing Outlook, 35(1), 34-37.

Bondy, K.N. (1983) 'Criterion-Referenced Definitions for Rating Scales in Clinical Evaluation', Journal of Nursing Education, 22(9), 376-382.

Bondy, K.N. (1984) 'Clinical Evaluation of Student Performance: The Effects of Criteria on Accuracy and Reliability, Research in Nursing and Health', 7, 25-33.

Boyatzis, R.E. (1982) The Competent Manager. New York: John Wiley & Sons.

Broadfoot, P. (1992) 'Changing Priorities in Assessment: An International Reform Agenda'. Keynote address to the NZ conference Qualifications for the 21st. Century, New Zealand Qualifications Authority, January 1992.

Burg, F.D., Lloyd, J.S. & Templeton, B. (1982) 'Competence in Medicine', Medical Teacher, 4(2), 60-64.

Burke, J.W. (1989) Competency Based Education and Training. London: The Falmer Press.

Butler, J.E. (1990) The Identification and Assessment of Competencies: The Nursing Project and its Implications. National Office of Overseas Skills Recognition Research Paper No. 4, DEET. Canberra: Australian Government Publishing Service.

Butler, J. (1992) 'Teacher Professional Development: An Australian Case Study', Journal of Education for Teaching, 18(3), 221-38.

Butler, J. (1994) 'From Action to Thought: The Fulfilment of Human Potential'. In Edwards J. (Ed.) Thinking: International

Interdisciplinary Perspectives. Melbourne: Hawker Brownlow.

Cason, G.J. & Cason, C.L. (1984) 'A Determinstic Theory of Clinical Performance Rating: Promising Early Result', Evaluation & The Health Professions, 7(2), 221- 247.

Conn, H.L. & Cody, R.P. (1989) 'Results of the Second Clinical Skills Assessment Examination of the ECFNG', Academic Medicine, Aug., 448 -453.

Dawkins, The Hon J.S., M.P (1989) Migrant Skills: Improving Recognition Processes. Canberra: Australian Government Publishing Service.

Eisner, E. (1993) 'Reshaping Assessment in Education: Some Criteria in Search of Practice', Curriculum Studies, 25(3), 219-33.

Ewell, P.T. (1991) 'To Capture the Ineffable: New Forms of Assessment in Higher Education', Review of Research in Education, 17, 75-125.

Fahy, K & Lumby, J. (1988) 'Clinical Assessment in a College Program', The Australian Journal of Advanced Nursing, 6(4), 5-9.

Fenton, M.V. (1985) 'Identifying Competencies of Clinical Nurse Specialists', The Journal of Nursing Administration, 15(12), 31-37.

Fitzpatrick, AR. (1989) 'Social Influences in Standard Setting: The Effects of Social Interaction on Group Judgements', Review of Educational Research, 59(3), 315 -328.

Flanagan, J.C. (1954) 'The Critical Incident Technique', Psychological Bulletin, 51, 327- 358.

Gonczi, A., Hager, P. & Oliver, L. (1990) Establishing Competency-Based Standards in the Professions. National Office of Overseas Skills Recognition Research Paper No. 1, DEET. Canberra: Australian Government Publishing Service.

Gonczi, A., Hager, P. and Athanasou, J. (1993) The Development of Competency-Based Assessment Strategies for the Professions. National Office of Overseas Skills Recognition Research Paper No. 8, DEET. Canberra: Australian Government Publishing Service.

Hager, P. (1994) 'Is There a Cogent Philosophical Argument Against Competency Standards?', Australian Journal of Education, 38,(1), 3-18.

Hager, P. and Butler, J. (1994) 'Problem Based Learning and Paradigms of Assessment'. In Chen, S.E., Cowdroy, R.M., Kingsland, A.J. and Ostwald, M.J. (Eds.) Reflections on Problem Based Learning. Sydney: Australian PBL Network.

Hager, P., Gonczi, A. and Athanasou, J. (1994) 'General Issues About Assessment of Competence', Assessment & Evaluation in Higher Education, 19(1), 3-16.

Haney, W., Madaus, G. & Kreitzer, A. (1987) 'Charms Talismatic: Testing

Teachers for the Improvement of American Education', Review of Research in Education, 14, 169-238.

Hayes, A. (1990) 'Developmental Psychology, Education and the Need to Move Beyond Typological Thinking', Australian Journal of Education, 34(3), 235-41.

Kermode, S. (1987) 'Pre-clinical Preparation of Undergraduate Nursing Students', Australian Journal of Advanced Nursing, 5(1), 5-10.

Klemp, G.O. (1980) The Assessment of Occupational Competence. Washington, DC: Report to the National Institute of Education.

Lareau, A. (1986) 'A Comparison of Professional Examinations in Seven Fields: Implications for the Teaching Field', The Elementary School Journal, 86(4), 553-569.

Maatsch, J.L. Huang, R.R., Downing, S.M. & Munger, B.S. (1987) 'Examiner Assessments of Clinical Performance: What do they tell us about clinical competence?', Evaluation and Program Planning, 10, 13-17.

McGaghie, W.C. (1989) Assessing Readiness for Professional Practice. An invited address to the American Educational Research Association. San Francisco, CA.

McDowell, B.J., Nardini, D.L., Negley, S.A. & White, J.E. (1984) 'Evaluating Clinical Performance Using Simulated Patients', Journal of Nursing Education, 23(1), 37-39.

Mansfield, B. (1989) 'Competence and Standards'. In J.W. Burke (Ed.) Competency Based Education and Training. London: The Falmer Press.

Masters, G.N. & McCurry, D. (1990) Competency-Based Assessment in the Professions. National Office of Overseas Skills Recognition, Research Paper No. 2. Canberra: Australian Government Publishing Service.

Menges, R.J. (1975) 'Assessing Readiness for Professional Practice', Review of Educational Research, 45(2), 173 -207.

Morgan, A. (1989) 'Interactive Assessment in a Nursing Program', Australian Journal of Advanced Nursing, 6 (4), 29-32.

Norman, G.R., Van Der Vleuten, C.P.M. and De Graaf, E. (1991) 'Pitfalls in the Pursuit of Objectivity: Issues of Validity, Efficiency and Acceptability', Medical Education, 25, 119-126.

Nursing Competencies Assessment Project Final Report (1990) Volume 1: Project Report; Volume 2: Assessment Technology; Volume 3: Data and

Archives. Adelaide: Nurses Board of South Australia.

Ross, M., Carroll, G., Knight, J., Chamberlain, M., Fothergill-Bourbonnais, F. & Linton, J. (1988) 'Using the OSCE to Measure clinical Skills Performance in Nursing', Journal of Advanced Nursing, 13, 45 -56.

Sadler, D.R. (1987) 'Specifying & Promulgating Achievement Standards', Oxford Review of Education, 13(2), 191-208.

Salman, R.D. (1981) 'Putting Testers to the Test: Relating Professional Licensing Examinations to Practice'. Conference on the Professions: Professional Education, Professional Examinations and Practice - Defining the Right Relationships, New York.

Samson, G.E., Graue, M.E., Weinstein, T. & Walburg, H.J. (1984) 'Academic and Occupational Performance: A Quantitative Synthesis', American Educational Research Journal, 21, 311 - 321.

Shulman, L.S. (1987) 'Assessment for Teaching: An Initiative for the Profession', Phi Delta Kappan, 69(1), 38-44.

Simons, H. and Elliott, J. (Eds.) (1989) Rethinking Appraisal and Assessment. Milton Keynes: Open University Press.

Spehn, M.R. (1985) 'The search for consultant competencies'. In D.D. Warrick (ed). Contemporary Organisation Development: Current Thinking and Applications. Glenview, IL: Scott, Foresman and Company, pp.402 -413.

Spencer, L.M. (1983) Soft Skill Competencies. Edinburgh: Scottish Council for Research in Education.

Sredl, H.J. & Rothwell, W.J. (1987) The ASTD Reference Guide to Professional Training Roles & Competencies. Vol 1. pp. 69-98. Amherst, MASS: Human Resource Development Press.

Sutnick, A.I., Barrows, H.S., Hart, I.R., Klass, D.J., Norcini, J.J. & Stillman, P.L (1989) Educational Commission for Foreign Medical Graduates, Clinical Skills Test Development Committee. Unpublished Report, USA.

Wandelt, M. & Slater Stewart, D. (1975) Slater Nursing Competencies Rating Scale. New York: Appleton-Century-Crofts.

Williams, L. (1989) Report of the Research Project to Develop Role Statements for Nurses. Brisbane: Board of Nursing Studies.

Wolf, D., Bixby, J., Glenn III, J. and Gardner, H. (1991) 'To Use Their Minds Well: Investigating New Forms of Student Assessment', Review of Research in Education, 17, 31-74.

Yates, G. and Chandler, M. (1991) 'The Cognitive Psychology of Knowledge: Basic Research Findings and Educational Implications', Australian Journal of Education, 35(2), 131-53.

1 Although paradigm is a term with a variety of meanings, we use it here to designate a taken-for-granted set of assumptions about what is involved in educational assessment.
2 The choice of the term 'judgement paradigm' is not intended to suggest that judgement is absent from assessment in the scientific measurement paradigm. Rather the term emphasises the point that professional judgement is at the heart of assessment according to the newer paradigm. This point is discussed in more detail later.
3 For other perspectives on challenges to traditional approaches to curriculum and assessment see Broadfoot 1992 and Hager 1994.
4 For more on the principles and practice of the judgemental model of assessment see Gonczi, Hager and Athanasou 1993 and Hager, Gonczi and Athanasou 1994.