

D I S C U S S I O N P A P E R

Teacher Observation in Student Assessment

Prepared for the Queensland School Curriculum Council by

Graham S Maxwell

School of Education, The University of Queensland



QUEENSLAND
SCHOOL
CURRICULUM
COUNCIL

Discussion Papers on Assessment and Reporting

Discussion Paper Number 1

Common and Different Features of Council and Board Approaches to Assessment and Reporting by Graham Maxwell, October 2001

Discussion Paper Number 2

Teacher Observation in Student Assessment by Graham Maxwell, October 2001

ISBN 0 7345 2320 3

© The State of Queensland (The Office of the Queensland School Curriculum Council), October 2001

Queensland School Curriculum Council
Level 27, 239 George Street
Brisbane, Queensland, Australia

PO Box 317, Brisbane Albert Street, Q 4002

Tel: (07) 3237 0794

Fax: (07) 3237 1285

Email: inquiries@qscq.qld.edu.au

Website: <http://www.qscq.qld.edu.au>

PREFACE

This discussion paper is part of a series being published and disseminated by the Office of the Queensland School Curriculum Council. The purpose of this series is to encourage discussion on various issues concerning assessment and reporting.

Teacher observation is one of several types of assessment techniques recommended by the Council in its *Position and Guidelines on Assessment and Reporting for Years 1 to 10* and in its syllabus documents for the key learning areas. Other assessment techniques include consultation, focused analysis, peer assessment and self-assessment. The *Position and Guidelines* state that: 'Observation involves teachers in observing students as they participate in planned activities. Teacher observation occurs continually as a natural part of the learning and teaching process and can be used to gather a broad range of information about students' demonstrations of learning outcomes' (p. 16).

Teacher observation has been valued as an important assessment technique in the lower primary school, but has in the past received less attention in later year levels. Secondary schools especially have favoured formal testing and structured assessment tasks rather than in-situ observation, although there has been some movement towards in-situ observation in some subjects in recent years, even in the senior years.

The Council's *Position and Guidelines* and syllabus documents elevate teacher observation to a more prominent position in the range of assessment techniques that teachers might use. Other techniques have been identified as consultation and focused analysis as well as peer and self-assessment. Focused analysis includes more formal assessment procedures such as set tests and set tasks. However, these categories of assessment techniques are not necessarily distinct. For example, as this paper explains, observation may be employed in association with focused analysis, especially where what is observed is student performance on a set task.

This discussion paper is not an official policy statement of the Council. Rather, it represents the views of the author, Dr Graham Maxwell, of the School of Education, The University of Queensland. As such, it offers a personal perspective on the issues. Dr Maxwell has been involved in research and consultation on assessment for many years in Australia, USA and UK, ranging over all sectors and levels of education. He has taught courses and conducted workshops on assessment for pre-service and in-service teachers for 30 years. He has also been involved in recent Council deliberations on assessment and reporting.

The audience for this discussion paper is professional educators, especially schoolteachers and administrators who must deal with assessment and reporting practice in classrooms and schools. Such people already know a great deal about assessment and reporting theories and practices, and this discussion paper builds on that knowledge. The hope is that the discussion paper will serve as a basis for professional debate, development workshops and collaborative planning.

The Office of the Council would be pleased to have your reactions to and comments on the discussion paper, as well as any examples of ways in which the discussion paper may have helped to clarify or resolve any theoretical or practical problems you are facing in the implementation of Council syllabuses. These reactions, comments and examples would assist the Office of the Council in deciding what further assistance it might be able to provide on these issues.

JE Tunstall
Director
Queensland School Curriculum Council

October 2001

CONTENTS

Preface	iii
Introduction	1
Arguments for valuing teacher observation	2
Collection and recording of evidence	5
Planning for teacher observation	7
Factors affecting validity of teacher observations and what to do about them	9
1. Prejudgments and prejudices	10
2. Selective perception	11
3. Providing inadvertent clues	11
4. Inappropriate inference	12
5. Inconsistency	12
Conclusion	13
Some useful references on teacher observation	14

INTRODUCTION

Teacher observation is one of the assessment techniques recommended in the *Position and Guidelines on Assessment and Reporting for Years 1 to 10* of the Queensland School Curriculum Council (the Council). For the implementation of Council syllabuses, assessment is seen as involving a variety of school-based (teacher-enacted) techniques for profiling student demonstrations of learning outcomes. Teacher observation, that is, observation of a student by a teacher, is one of those techniques. It can be used in conjunction with other techniques.¹

Teacher observation has been accepted readily in the past as a legitimate source of information for recording and reporting student demonstrations of learning outcomes in early childhood education. As the student progresses to later years of schooling, less and less attention typically is given to teacher observation and more and more attention typically is given to formal assessment procedures involving required tests and tasks taken under explicit constraints of context and time. However, teacher observation is capable of providing substantial information on student demonstration of learning outcomes at all levels of education.

For teacher observation to contribute to valid judgments concerning student learning outcomes, evidence needs to be gathered and recorded systematically. Systematic gathering and recording of evidence requires preparation and foresight. This does not necessarily mean that all aspects of the process of observation need to be anticipated but that the approach taken is deliberate rather than happenstance. It is necessary, at least, to know in advance both what kinds of learning outcomes are anticipated and how evidence will be recorded. Adequate records are essential for good assessment.

Teacher observation can be characterised as two types: incidental and planned.

- *Incidental observation* occurs during the ongoing (deliberate) activities of teaching and learning and the interactions between teacher and students. In other words, an unplanned opportunity emerges, in the context of classroom activities, where the teacher observes some aspect of individual student learning. Whether incidental observation can be used as a basis for formal assessment and reporting may depend on the records that are kept.
- *Planned observation* involves deliberate planning of an opportunity for the teacher to observe specific learning outcomes. This planned opportunity may occur in the context of regular classroom activities or may occur through the setting of an assessment task (such as a practical or performance activity).²

¹ Other possibilities for collecting evidence by observation exist, including observations by another teacher, observations by other students, and student self-observation. This paper concentrates on teacher observation and does not consider these additional forms of observation. Ultimately, the teacher must judge the validity and relevance of all forms of evidence, so these alternative forms of evidence involve similar principles of assessment to those applicable to other forms of evidence, including teacher observation.

² 'Classroom' should be interpreted liberally. It may include settings outside the school, such as field excursions, public presentations and work experience. What matters is the presence of an assessor, typically the teacher, to observe and record the event.

ARGUMENTS FOR VALUING TEACHER OBSERVATION

Teacher observation is an important but underutilised assessment technique. It is sometimes argued that teachers are unable to make appropriate and dependable assessment judgments from observations of students in natural settings. The table (below) details some of the claims of this viewpoint, together with some relevant rebuttals. Handled carefully, teacher observations can provide important evidence for assessment judgments. In some cases, they provide the only way of obtaining evidence about particular learning outcomes, especially those involving practical techniques, performance activities, ‘real life’ projects and group work.

Table 1: Arguments against teacher observations and rebuttals of those arguments

Arguments against teacher observations	Rebuttal of those arguments
<p><i>Lack of representativeness</i> Students may not demonstrate all relevant learning outcomes in natural settings. They may know or know how but the context may not prompt them to demonstrate this.</p>	<p>Learning outcomes that have not been demonstrated can be deliberately prompted. Assessment should be planned as well as incidental. Teachers can ensure that assessment is comprehensive.</p>
<p><i>Lack of observation</i> Teachers may not observe the demonstration of a learning outcome when it occurs, either because their attention is elsewhere or because they fail to recognise it.</p>	<p>Over time, teachers have many opportunities for observation. It is not critical if particular opportunities for observation are missed. Some observation is deliberate and focused.</p>
<p><i>Lack of control of influences</i> The student can derive unintended cues and prompts from the setting, even from the teacher, and these can be unnoticed by the teacher. Student performance may then be misinterpreted.</p>	<p>No single occasion is sufficient for judging a student’s demonstration of learning outcomes. Multiple opportunities and a variety of contexts allow cross-checking the robustness of the student’s performance.</p>
<p><i>Lack of standardisation</i> All students do not undertake the same tasks under the same conditions. Teacher judgments of student demonstrations of learning outcomes are therefore undependable.</p>	<p>Quality requirements for teacher judgments are ‘evidence-based’ and ‘defensible’. Tailoring and adaptation allow optimum student performance and holistic interpretation of the evidence (taking contextual factors into consideration).</p>
<p><i>Lack of objectivity</i> Teacher judgments are subjective and prone to inconsistencies. Too much is left to the discretion of the teacher.</p>	<p>All assessment involves sequences of subjective decisions; mechanistic marking schemes reflect earlier design decisions. Procedures to strengthen and verify teacher judgments can be introduced.</p>
<p><i>Possibility of stereotyping</i> Subjective judgments allow the possibility of stereotyping of students in terms of other performances or characteristics.</p>	<p>Stereotyping is not inevitable. Each assessment occasion can be approached as a fresh opportunity to test hypotheses derived from prior impressions.</p>
<p><i>Possibility of bias</i> Subjective judgments allow the possibility of conscious or unconscious bias for or against particular individuals or groups.</p>	<p>Conscious bias is unethical. Unconscious bias requires constant vigilance. It is difficult for bias to survive evidence-based justification to students and their parents (a form of accountability).</p>

An important argument in support of teacher observation is that teachers have access to a rich and diverse range of evidence on student learning outcomes from observations of their students; and that the capability of teachers to collect and interpret this range of evidence should be respected. Otherwise, a rich source of evidence on student learning outcomes is being ignored. In any case, the issue is not whether teacher observations should be used — since they are necessarily used informally during teaching — but how teacher judgments can be strengthened and improved so that they can be used for formal purposes, especially for reporting and certification.

A strong justification for using teacher observation in assessment is its capacity to enhance assessment validity. By extending the range of possible assessments, teacher observation allows assessment to be more:

- *comprehensive* — ensuring recognition of all desired learning outcomes, especially those not otherwise assessable than in classroom contexts;
- *connected* — situated within familiar learning contexts and closely related to curriculum frameworks, learning experiences and pedagogical planning;
- *contextualised* — sensitive to the effects of context on performance and deriving assessment evidence from a variety of situations and occasions;
- *authentic* — interesting, challenging, worthwhile and meaningful to students;
- *holistic* — emphasising relatedness and connections in learning and involving performance on complex wholes rather than separate components.

All of these characteristics can be supported as important for high quality learning and assessment. Assessments with these characteristics have

better representation of,
clearer relevance to, and
stronger consequences for

desirable learning outcomes. Appropriate representation, relevance and consequences are often now recognised as the requirements of valid assessments.

Past practice in assessment, particularly in secondary schools, has not accorded much recognition to teacher observation for formal purposes. Some people have argued that while teacher observation is necessarily a component of classroom teaching, a clear distinction should be made between informal and formal assessment. For example, some educators argue that assessment situations should not be confused with learning activities.³ This draws too firm a distinction. Certainly, there should not be strong accountability for first attempts. Also, reports need to indicate the progress made by students at the time of reporting. However, a firm distinction between assessment situations and learning activities stems from over-

³ This is argued by Caroline Gipps in her book, *Beyond Testing: Towards a Theory of Educational Assessment*, The Falmer Press, London, 1994.

concern for comparison and ranking. Where the aim is to map the student's profile of demonstrated learning outcomes, standardised comparison is not the issue. The question is simply what justifiable evidence is there for concluding that the student has demonstrated a particular learning outcome. This does not require students to be 'tested' under controlled — and artificial — conditions.

An alternative vision is one where assessment becomes incidental to and indistinguishable from learning activities. This allows for the strongest connection between pedagogy and assessment and, as already argued, a strengthening of the quality of the learning and the validity of the assessment. It can be argued that unless there is a strong connection between pedagogy and assessment, the assessment will be disembodied and discriminatory, that is, unconnected to any means for improving student learning and privileging students with existing cultural capital. Such an approach focuses on the student's *best* performance over time and values the *progress* they are making (similar to 'personal best' in athletics).

Howard Gardner puts it this way: 'Rather than being imposed "externally" at odd times during the year, assessment ought to become part of the natural learning environment. As much as possible it should occur "on the fly", as part of an individual's natural engagement in a learning situation. Initially, the assessment would have to be introduced explicitly; but after a while, much assessment would occur naturally on the part of student and teacher, with little need for explicit recognition or labelling on anyone's part. ... As assessment gradually becomes part of the landscape, it no longer needs to be set off from the rest of classroom activity. As in a good apprenticeship, the teachers and the students are always assessing. There is also no need to "teach for the assessment" because the assessment is ubiquitous; indeed, the need for formal tests might atrophy altogether.'⁴

⁴'Assessment in context: The alternative to standardized testing', in B.R. Gifford & M.C. O'Connor (Eds), *Changing assessments: Alternative views of aptitude, achievement and instruction* (pp. 78-119), London, Kluwer, 1992.

COLLECTION AND RECORDING OF EVIDENCE

All assessment requires the collection and recording of evidence of student learning. For the implementation of Council syllabuses, it has been recommended that the evidence focus on the demonstration of learning outcomes. Evidence is documentation that records, illustrates or confirms student demonstrations of learning outcomes.

Collection and recording of evidence is necessary for two reasons:

- accountability — justification of the assessment judgments; and
- verification — confirmation of the assessment judgments.

Accountability and verification are key factors in assuring the quality of assessments.

- *Accountability* (justification) means being able to explain and defend assessment judgments to students, their parent(s) and other teachers.
- *Verification* (confirmation) means being able to revisit the foundations for assessment judgments — being able to check their completeness, relevance and veracity.

Teacher observations are primarily directed at the observation of events, performances and activities. In some cases, an artefact may be produced as a consequence of the event, performance or activity. In other cases, no artefact is produced and the event, performance or process itself is the sole focus of attention. An artefact is something constructed by the student, for example, a worksheet, a piece of writing, a design, a painting, a composition, a webpage — in other words, a product of some kind. Teacher observation is not primarily concerned with the artefact itself but with the way in which the artefact was produced, that is, with the process.⁵

Evidence of process, whether or not there is a resultant artefact, may involve either direct record or written record. These two types of record have different characteristics.

A *direct record* keeps a ‘trace’ of the event through an audio-recording, a video-recording or a sequence of photographs. The activity or event might be, for example, a speech, a dramatic presentation, a group activity or a practical task. The term ‘trace’ emphasises that the record is not the same as the event itself. At best, it allows some features of the event to be represented and recalled. Some features of the event may be lost, such as the ‘feel’ of the occasion or the ‘spark’ between presenter and audience. Some features of the event may be filtered or distorted by the medium of recording, for example, through positioning and handling of the

⁵ In some cases, a sequence of artefacts may be produced and these may provide a progressive record of stages of production. These could involve, for example, a sequence of written drafts, initial designs, trial compositions, or tentative frameworks. In this case, the artefacts indicate milestones of development towards the final product. It is important to keep the total sequence of artefacts, together with annotations about contextual factors, such as the way in which the student has made use of comments and suggestions, so that a complete interpretation of what the student has done can be made.

recording device. It is important, therefore, to realise that such a record offers only partial representation of the event. Nevertheless, such ‘traces’ are better than having no record at all.

A *written record* can take the form of an observation sheet or a logbook (diary of events). Observation sheets can be more or less structured: at one extreme they contain checklists of learning outcomes; at the other extreme they contain broad categories for writing on-the-spot comments or annotations; and in between these extremes is a combination of them both.⁶ A logbook provides a record of critical incidents or key comments (sometimes referred to as an anecdotal record); for accuracy of recall, entries need to be made as soon after the event as possible. The student’s name and the date also need to be clearly recorded. A sequential collection of such records is sometimes called a ‘running record’.

When keeping an observation sheet or a logbook, written entries can be (relatively) high inference or (relatively) low inference. ‘High inference’ means that a judgment or interpretation is made, whereas ‘low inference’ means that the specifics of the event are described (without any attempt to interpret what they signify). Thus, using a *checklist of core learning outcomes* would involve high inference, whereas providing a *descriptive account of student performance* without direct reference to core learning outcomes could involve low inference. In the latter case, judgments relating to the demonstration of learning outcomes can be delayed until a variety of evidence has been collected.

The advantage of low inference observations is that they are more ‘objective’ or ‘transparent’ and can provide a ‘closer-to-the-event’ basis for later verification; the specifics of the event are more easily accessible. The advantage of high inference observations is greater ease and efficiency in record keeping, but the specifics of the event are not then retrievable from the record. It is possible to record both a judgment (high inference) and a description (low inference), thus retaining some of the benefits of each.

Descriptive accounts can include written commentary on student performance under specific headings on an observation sheet; critical incidents or significant events recorded in a logbook; and key performance features recorded in a logbook. Critical incidents and significant events are particularly noteworthy instances of the demonstration of particular learning outcomes (or the lack of it), especially those observed for the first time or demonstrated in a particularly dramatic or unexpected way. Key performance features are salient features of an observed performance whether or not they are particularly ‘critical’, including evidence that strengthens or confirms early judgments relating to student demonstrations of learning outcomes.

When an artefact, direct record or descriptive account is kept and placed in the student’s assessment portfolio, the artefact, direct record or descriptive account can be referred to again at a later time in order to retrieve the specifics of student performance. This allows assessment judgments to be delayed until a convenient time or allows assessment judgments to be

⁶ Rating scales are not mentioned here since arbitrary distinctions of quality are not part of the assessment position espoused by the Council. However, the sequences of core learning outcomes (arranged in levels along a developmental continuum) provide quasi-rating scales. It is a matter of convenience and style whether a sequence of core learning outcomes is characterised as a quasi-rating scale (showing where the student is positioned along the developmental continuum) or simply an expanded checklist (showing which core learning outcomes the student has demonstrated).

verified on a subsequent occasion, for example, for purposes of moderation. It is difficult to verify written records of judgments without an accompanying artefact, direct record or descriptive account as a reference point.⁷

Table 2: Summary of types of teacher observation evidence

<p>Focus on product — keep artefact(s)</p> <p>Focus on activity — record process</p> <ul style="list-style-type: none"> • Direct record ('trace') <ul style="list-style-type: none"> ○ Audio-tape ○ Video-tape ○ Photographs • Written record <ul style="list-style-type: none"> ○ Observation sheet <ul style="list-style-type: none"> ▪ Checklist (high inference) ▪ Description (low inference) ○ Logbook <ul style="list-style-type: none"> ▪ Description of critical incidents ▪ Description of key performance features
--

PLANNING FOR TEACHER OBSERVATION

Teacher observations cannot be useful without planning. Different types of evidence require different types of planning.

An essential requirement for all types of evidence is anticipating the kinds of learning outcomes that may be demonstrated. This is particularly important where observation is incidental and where judgments (rather than descriptions) are recorded. Council syllabuses provide a framework of learning outcomes that serve as the perceptual reference points for recognising the characteristics of student performance. The framework of learning outcomes makes available to the teacher concepts and language for recognising and describing what a student knows and can do. Learning the structure, language and concepts of the framework therefore is a key aspect of planning for teacher observation, as it is too for teaching.

Incidental observation necessarily involves little additional planning, apart from the normal planning of classroom learning activities for students. Incidental observation is opportunistic,

⁷ Where there is no supporting evidential record for the judgment, verification of the judgment is strictly impossible. The record of judgment needs to be considered in the context of other evidence collected from different times and events. Where it corroborates other evidence, the judgment is strengthened. Where it contradicts other evidence, more evidence may be needed. Ultimately, the weight of evidence is what matters. 'Other evidence' could include the judgments of other observers, that is, other records of judgment of the same event.

capitalising on revelations of student learning during regular classroom learning activities. In this sense it cannot be planned. It is essentially unanticipated. It can only be recorded through descriptions in a logbook. Although there may sometimes be an artefact to provide corroboration for the teacher's observation, any process details depend on teacher description. Incidental observation is therefore the weakest form of teacher observation and would preferably be used only as supplementary evidence to support other forms of evidence. Relying on incidental observation alone would be unsatisfactory (see caveats below).

Planned observation can involve planning for 'in situ' observation (in learning situations) or planning for set assessment tasks. There is little to distinguish these two situations in practical terms. However, as assessment becomes more important, particularly in Years 8 to 10, students may need to know when they are being assessed, since they may otherwise choose not to show their actual capabilities. Absence of demonstration of learning outcomes might not indicate incapability of demonstrating those learning outcomes but lack of appropriate challenge or opportunity. Formal assessment occasions would appear to become more important in the secondary school than in the primary school, at least for the present.⁸

For all planned observations, whether 'in situ' or set tasks, thought needs to be given to how the event and/or the observations will be recorded. Consideration needs to be given to whether a direct record will be kept and what form of observation record will be made. The validity of teacher observations is strengthened by preparing an observation sheet that allows systematic recording of observations and judgments. An observation sheet may include checklists of learning outcomes and/or categories for describing student activities and performances. Learning outcomes might be made more explicit by listing their elaborations, components or criteria, that is, by providing more detail on the characteristics of the desired learning outcome.

The advantages of prepared observation sheets include:

- opportunity to share learning expectations with students in advance
- encouragement of student self-monitoring and self-assessment
- clarification of the desired learning outcomes to guide learning
- focus on the desired learning outcomes to guide teaching
- cuing of attention to the full range of relevant learning outcomes
- having available an explicit and standard recording format
- ease of recording of student performance characteristics
- structured means of providing feedback to students.

⁸ In the long term, taking up Howard Gardner's vision (see footnote 4), a more natural approach to assessment would require that classrooms become more like normal work environments. This does not necessarily mean that the student (as the worker) is under constant surveillance but that there are opportunities to demonstrate their capabilities in situations that really matter (but are preferably 'authentic' rather than 'artificial'). The necessary shift in assessment practice to support this is from 'one-off testing' to progressive demonstration of 'best (sustained) performance'. Reformed assessment practice would place more onus on the student to demonstrate the desired learning outcomes (and to indicate when they think they are doing so) together with specific 'invitations' to students to demonstrate their level of development in relation to sequences of learning outcomes of increasing complexity.

Disadvantages of prepared observation sheets include:

- the need to allow for several levels of learning outcomes on a single sheet
- it can be difficult to anticipate all the learning outcomes that might appear
- it is possible that other serendipitous learning outcomes will be missed
- students' learning may be constrained by listed learning outcomes.

The disadvantages are outweighed by the advantages. They can be overcome, in any case, by careful design of the observation sheet, tailoring it to the current stage of student development, and allowing space for additional observations to be recorded. Observation sheets should be used as a tentative organising structure for recording teacher observations rather than a limiting framework for the actual observations.

Space also needs to be provided on the observation sheet for including descriptive details of the context. These details need to include any characteristics of the setting or the occasion that could have influenced the student's performance, either positively or negatively, and that might be relevant in making a judgment about whether the student has demonstrated particular learning outcomes. The details can be physical (e.g., uncomfortable surroundings), psychological (e.g., personal attributes in stressful situations) or social (e.g., other events in the life of the school or the student).

Through all of this, it must be remembered that any written record of observations is necessarily selective. Only certain features of student performance are likely to be noticed and can be recorded. Therefore, having a clear understanding and ready access to the framework of expected learning outcomes is essential. One technique for reducing the cognitive demands of open observation is 'spotlighting'. This means targeting specific learning outcomes (across several levels of a strand) on particular occasions. This has the added advantage of ensuring systematic coverage of all relevant learning outcomes. However, it should not be pursued so religiously that evidence of other learning outcomes outside the spotlighting target is ignored.⁹

FACTORS AFFECTING VALIDITY OF TEACHER OBSERVATIONS AND WHAT TO DO ABOUT THEM

Teacher observations will be valid to the extent that the evidence is appropriately recorded and interpreted, that is, whether:

- the recorded evidence accurately represents the observed student performance
- the interpretation (judgment) of this evidence is justifiable.

Accurate recording requires transparent and unbiased perception of the student's performance. Justifiable interpretation requires careful consideration of what the student's performance signifies, in terms of learning outcomes, taking into consideration any factors

⁹ Margaret Forster and Geoff Masters discuss 'spotlighting' in *Performances: Assessment resource kit*, Camberwell, Victoria, Australian Council for Educational Research, 1996.

that may have influenced the performance. The use of the term ‘justifiable’ here emphasises that there may not be a single unequivocal interpretation of the evidence but rather that the interpretation should withstand challenge as being reasonable and defensible. There also may be a requirement that the interpretation be consistent with the interpretations of other teachers.¹⁰

The following discussion covers some factors that can affect the accuracy of the recorded evidence or the justifiability of the interpretation or both. These factors are caveats or warnings about the need for careful deliberation and for using multiple sources of evidence in making judgments about demonstrations of learning outcomes.

1. Prejudgments and prejudices

Prejudgments are judgments formed in advance on the basis of prior information or initial impressions. Prior information may come from other teachers or from the teacher’s familiarity with the student’s performance on previous occasions or in different contexts. The information may be irrelevant to the judgment being made (about performance on this occasion) but influences that judgment, in some cases to such an extent that the teacher does not actually ‘see’ the true nature of the student’s current performance. That is, the expectation of what the student’s performance will be like overrides the perception of what it is really like. This is sometimes referred to as the ‘halo effect’, though a better term would be the ‘masking effect’. The effect can operate in the ‘positive’ (or ‘halo’) direction, causing the teacher to think that a student has demonstrated a learning outcome that they have not. But it can also operate in the negative direction, causing the teacher to think that a student has not demonstrated a learning outcome when in fact they have.

Prejudices are personal theories or attitudes presuming that students with particular characteristics will perform in certain ways. That is, students may be stereotyped through their membership of certain groups, particularly groups defined by social characteristics, culture or gender. The effect is again to mask the true nature of the student’s performance, this time on the basis of an expectation relating to the group. This expectation of the group may itself be inaccurate. That is, the ‘prejudice’ may be towards the group, not just the individual. Again, this effect can operate in either direction: judging that a learning outcome has been demonstrated when it has not been demonstrated or judging that a learning outcome has not been demonstrated when it has.

Steps can be taken to guard against judgments and prejudices. The following five approaches should prove useful.

(a) *Adopt a hypothesis-testing stance* towards all information and evidence. That is, prior expectations should be treated as tentative possibilities to be confirmed or amended in the light of the evidence. Judgments are not foreclosed before all the evidence is available. The questions to ask are: ‘Have I seen this evidence for what it truly signifies?’ and ‘Is my perception of the evidence unclouded by extraneous factors and preconceptions?’

¹⁰ Approaches to moderation of assessments among teachers are discussed in another discussion paper.

(b) *Use low-inference (descriptive) approaches* for recording evidence and for initial interpretations of the evidence. Being descriptive in the first instance ensures that the focus is on what the student actually said or did, thus making it less likely that the actual features of the student's performance will be overlooked or distorted.

(c) *Don't depend exclusively on observation* for evidence of demonstration of learning outcomes. Use other forms of evidence as cross-checks.

(d) *Build up a variety of evidence over time* until there is some confidence in the judgment that particular learning outcomes have been demonstrated.

(e) *Where possible, have your observations verified by another observer*, for example, another teacher. Talk about any differences in perception or interpretation with a view to understanding why they occur and developing sensitivity to any personal prejudgments or prejudices.

2. Selective perception

Selective perception involves seeing and hearing what we are predisposed to see and hear. This is not a conscious choice. Rather, it results from our intuitive expectations and psychological preconceptions. Psychological research shows that sometimes these can produce very powerful distortions of our perceptions. For example, we can so strongly anticipate that the student will say or do something that we imagine that they have said or done it when in fact they have said or done something quite different. Or we can so strongly anticipate that the student will not be able to say or do something that we fail to notice when they do. Steps that can be taken to guard against selective perception are the same as those for preconceptions and prejudices (see above).

3. Providing inadvertent clues

When the focus of an observation is a verbal interchange between the student and the teacher, especially a question-answer sequence, it is possible for the teacher to unwittingly provide clues to the student and to draw the wrong conclusions from the student's answers. This can occur through providing more and more explanation or reframing of the question with commensurate reduction in the range and complexity of potential answers. Provided this is done consciously and deliberately, such 'scaffolding' can be an excellent technique for establishing the limits of the student's knowledge and capability (the student's 'zone of proximal development').

If a verbatim record of the interchange is kept, it is possible to reanalyse and reinterpret the student's response. If a verbatim record is not kept, it is important for the teacher to be sensitive to the possibility of providing unintended clues and to seek confirmatory evidence on other occasions and in other ways.

4. Inappropriate inference

Inappropriate inference means drawing the wrong conclusions. This can occur through focusing on the wrong features of the student's performance or through categorising the performance under the wrong learning outcomes.

Some examples of focusing on the wrong features of student performance are:

- accepting simple/closed answers as if they were open/complex answers
- interpreting non-verbal expressions (gestures, attention, eagerness) as indicating verbal proficiency
- confusing effort and involvement for performance and achievement.

In these cases, not only is an inappropriate judgment being made concerning what learning outcomes have been demonstrated, but the limits of the student's knowledge and capability have not been fully explored (and challenged). The student has been short-changed.

As with any written statement of a standard against which to judge performance, learning outcomes admit a range of interpretations. Common interpretations need to arise out of sharing of understandings and this may take some time (and perhaps some form of moderation). An aspect of interpretation is matching a performance to a learning outcome, that is, what are and what are not appropriate demonstrations of the learning outcome. While there will always be some variability among teachers in deciding which learning outcomes are demonstrated by particular performances, some matches can be identified as mistaken. These are cases where there can be no logical basis for concluding that the characteristics of the performance match the learning outcome. Care needs to be exercised to reduce the likelihood of such cases.

Inappropriate inference occurs during the judgment phase of assessment, that is, what is to be concluded from the evidence. The conclusion might be that more, or different, evidence is needed. Ensuring appropriate inference requires constant vigilance and clarity of understanding of the learning outcomes. Some useful questions are:

- Which learning outcomes am I spotlighting to gather evidence?
- Am I focusing on the appropriate features of the student's performance?
- Have I missed any important features of the student's performance?
- Have I matched the evidence to the appropriate learning outcomes?
- Am I clear about what are and are not appropriate demonstrations of these learning outcomes?
- Can I defend my judgment by explaining how the evidence indicates demonstration of the particular learning outcomes?
- Can I confirm my judgments by sharing them with another teacher?

5. Inconsistency

Inconsistency in assessment judgments means that evidence is interpreted differently in different circumstances. These different circumstances can include different occasions,

outcomes or students (for the same teacher), different teachers (within the same school) or different schools.

Inconsistency across different occasions, outcomes or students means that the teacher interprets the same (or similar) evidence differently depending on when the evidence was collected (and interpreted), what the evidence was about (that is, the learning outcomes involved) and who was the focus of the assessment (that is, different students treated differently). Such inconsistencies derive from errors of prejudgment or prejudice, selective perception, inadvertent clues, or inappropriate inference. Improvements in consistency therefore depend on taking action to guard against such errors. It is also useful to make spot checks to see whether similar evidence (from different occasions, relating to different outcomes, and for different students) is being interpreted consistently. This involves placing two or more pieces of evidence side by side and asking whether they are in fact indicative of the same (or different) learning outcomes.

Provided teachers are consistent within themselves, inconsistency between teachers (in the same school and in different schools) results from a lack of common understanding of core and discretionary learning outcomes and of what are appropriate demonstrations of those learning outcomes. The QSCC Research Report *Consistency of Teacher Judgment* has already explored this issue in detail and provided some guidelines for improving consistency between teachers.¹¹ In relation to teacher observation, these guidelines can be reframed to include the following: collaborative planning for classroom observation; developing common tasks and activities; developing common recording devices; comparing samples of student performance (moderation); sharing understandings of the developmental sequence of learning outcomes; sharing understandings of core and discretionary learning outcomes; sharing understandings and practices in teacher observation; and assessing collaboratively (through joint or swapped observation in each other's classrooms or through joint reviewing of records of classroom performances).

CONCLUSION

This discussion paper sets out a justification for using classroom observation as an important part of assessment strategies in assessment undertaken by teachers. It also highlights the importance of ensuring that adequate evidence is collected in order to support the need for verification and accountability of assessment judgments. The purpose of the paper is to stimulate consideration of the role that observation can play in assessment and to extend the range of possibilities for collecting evidence relating to the demonstration of learning outcomes to include incidental observation in classrooms. A framework for thinking about different types of observation and different types of evidence has been provided, as well as some cautions about possible sources of inconsistency in teacher judgment and how these might be reduced. What is needed now is for teachers to try various methods of using teacher observation as part of their assessment program and to share these with other teachers.

¹¹ Office of the Queensland School Curriculum Council, *Consistency of Teacher Judgment: Research Report*, Brisbane, QSCC, 2000.

SOME USEFUL REFERENCES ON TEACHER OBSERVATION

Airasian, P. W., 1994, Sizing-up assessment, in *Classroom assessment* (2nd ed.) (pp. 29-61), McGraw Hill, New York.

Assessment and Reporting Directorate, 1997, *Strategies for assessment and reporting: Primary schools*, Department of School Education, Ryde, NSW.

Brady, L. & Kennedy, K., 2001, *Celebrating student achievement: Assessment and reporting*, Prentice Hall, French's Forest, NSW.

Forster, M. & Masters, G., 1996, *Assessment resource kit: Performances*, Australian Council for Educational Research, Camberwell, Vic.

Genesee, F. & Upshur, J.A., 1996, *Classroom-based evaluation in second language education*, Cambridge University Press, Cambridge, UK.

Henning-Stout, M., 1994, *Responsive assessment: A new way of thinking about learning*, Jossey-Bass, San Francisco.

Hills, T., 1992, Reaching potentials through appropriate assessment, in S. Bredekamp & T. Rosegrant (Eds), *Reaching potentials: Appropriate curriculum and assessment for young children (Vol. 1)* (pp. 43-63), National Association for the Education of Young People, Washington, DC.

Masters, G. & Forster, M., 1996, *Assessment resource kit: Developmental assessment*, Australian Council for Educational Research, Camberwell, Vic.