Assessment of higher order thinking skills: A discussion of the data from the 2001 random sampling exercise and a workshop for teachers

Erica Bell, Reg Allen, Pam Brennan

Queensland
Board of Senior Secondary School Studies

Floor 7, 295 Ann St Brisbane Queensland

PO Box 307 Spring Hill Qld 4004

Telephone: (07) 3864 0299

Fax: (07) 3221 2553

Email: office@qbssss.edu.au

Website: www.qbssss.edu.au

(job 1372)

CONTENTS

ACKNOWLEDGMENTS

The research team would like to acknowledge the contribution of the Board's panels who provided the data discussed in the main body of this report. Their continuing contribution to building high-quality assessment practices in senior studies is highly valued by the staff of this Office.

Thanks are also owed to the review officers who provided many suggestions and insights that helped shape the report, and to Peter Jordan (Manager, Assessment and Panels) who contributed his particular expertise in this area.

Mr Chris Blood (Head of Mathematics, Brisbane Boys College) also provided helpful feedback on the report.

Erica Bell (Manager, Policy & Evaluation) designed the project and wrote the main report. Reg Allen (Deputy Director, Testing & Analyses) and Pam Brennan (Senior Research Officer) wrote the workshop that appears in appendix two of this report, redrafting and updating a set of workshop materials developed in 1996 by Erica Bell and Reg Allen. Marg Veerman (Senior Research Officer) provided the annotated list of resources that appears in appendix two.

FOREWORD

Following collection of data about higher order thinking skills as part of the Board's 2001 Random Sampling exercise, this discussion paper is written for all teachers interested in developing assessment of higher order thinking skills. This includes teachers who have received advice from district review panels and/or random sampling panels that there is scope for improving assessment of higher order thinking skills. It describes the data about higher order thinking obtained from the 2001 random sampling exercise, and explores what might be meant by key terms about higher order thinking used by these panels when they examined particular school submissions. Appendix one provides a workshop that schools can use to develop depth and quality in their assessment instruments, as well as the best possible match between intention and practice. Appendix two provides an annotated list of resource papers about higher order thinking.

This document is designed to work as one of a number of resources that schools can consult as part of their continuing effort to achieve excellence in assessing higher order thinking. It provides some generic understandings of higher order thinking skills and ways of developing assessment instruments, rather than entering the fine detail of subject-specific debates. Teachers who want specific advice about assessment of these skills in their subjects should, of course, consult the relevant review officer.

AIM

The study tries to go a little further than simply documenting the data obtained from 2001 random sampling panels reflecting on higher order thinking in school submissions of student work and assessment instruments.

The underlying questions motivating this paper could be posed as follows.

When district review panels provide advice to schools about problems with standards or assessment packages, they often refer to the need to improve assessment of higher order thinking skills. How can we report the data from the 2001 random sampling exercise in ways that will help schools reflect on the key terms and concepts used by panels commenting on higher order thinking? What practical suggestions can the Board provide to schools that want to develop their assessment practices in this area?

That is, this document does not aim simply to record the results of a data-gathering exercise or provide a discussion paper. We wanted to avoid writing a research paper that teachers might (or might not) read and say 'Yes, all of those data and discussions are interesting, but what can I do to develop the assessment instruments in my subject?'

The workshop materials in appendix one are designed to support schools to achieve greater depth in their assessment practices, and the best possible match of intention and practice. These materials are designed to be used in a whole school workshop situation, so that staff from different subject areas can pool their assessment experience and expertise. The workshop adopts a holistic approach (helping teachers examine the links between intention and practice) rather than offering a checklist approach. This is because in our trials of the workshop instrument it became clear that giving teachers a checklist of particular knowledge and skills that should be in assessment instruments (for example a checklist of CCEs that relate to higher order thinking, or any other checklist) had been tried before in these and other schools, with limited usefulness. What our trials suggested was that a checklist was well and good, but it was not the same as actually supporting teachers to examine and develop the best possible relationship between intention and practice in assessment. And it appears that in the area of higher order thinking the match or mismatch between intention and practice is crucial precisely because of the demands this end of assessment makes on assessment design expertise.

For teachers in our trial there was less interest in discussion of what assessment instruments ought to be achieving (e.g. discussion of a list of CCEs) and more interest in ensuring that the instruments students actually experienced did do, in practice, what they were intended to do. It is this that became the focus of our workshop materials.

METHOD

Two main strategies were used to answer the research question about how we can report the data from the 2001 random sampling exercise in ways that will be useful to schools that want to reflect on the key terms and concepts used by panels commenting on higher order thinking.

1. Identification of key terms and words associated with those key terms, using comments from district review panels that, in 2001, looked at a sample of work from 2000 as part of the random sampling exercise. In this year we collected data and comments specifically about higher order thinking skills as this was the 'random sampling' special topic.
2. Some discussion of some key terms and/or related concepts in the international literature. Clearly, the relevant panel is best placed to explain what their comments mean in the context of a particular assessment package; however, it is also true that the broad advice that panels often give suggests the importance of developing generic understandings of higher order thinking skills. The discussion of the literature is intended to offer a sample of some ideas in the literature (best fleshed out with broader reading, as indicated by the annotated list of resources in appendix two).

The question 'What practical suggestions can the Board provide to schools that want to develop their assessment practices in the area of higher order thinking?' can obviously be answered with many suggestions. Appendix one of this study provides some workshop materials that were trialled in seven schools (across Queensland, including more remote schools, and schools with particular issues in their QCS Test data such as lower than expected QCS Test performance when you consider their levels of achievement).

The rationale for this kind of workshop offering a 'whole school' approach to developing the match between intention and practice is given in the aims section. It represents an update of some Board workshop materials, *Studying Assessment Practices*, published in 1996 (Bell).

BACKGROUND: THE CONTEXT OF RANDOM SAMPLING

To understand what panels involved in the Board's 2001 random sampling exercise meant when they commented that in a particular school subject-group there is a need for development of assessment of higher order thinking skills, we need to first understand what random sampling involves.

Random sampling is one of the Board's quality-assurance procedures. It offers information about how successful the Board's procedures are in providing schools with suitable advice about standards, advice that leads to appropriate action by schools.

The Board seeks to maintain the statewide comparability of levels of achievement in each Board subject. It does this by providing advice to schools about the suitability of the standards that they apply. Panels review submissions and provide advice about the appropriateness of standards applied by schools on at least three occasions during the moderation process: in February at Year 11 monitoring, in October at Year 12 verification, and in March of the following year at random sampling.

The 2001 Random Sampling Project selected student folios from 20 subjects in schools where a minimum of 15 students completed the subject in Term 4 in 2000. For each school subject-group, the school was asked for a submission comprising: exit folios of nine designated students, all assessment instruments, a copy of the school work program, and a profile of each selected student's achievement (with the exit decisions masked on the profile).

For each subject, a maximum of 20 school submissions was sampled. School submissions were sent to a random sampling panel in another district. Each random sampling panel consists of two review panellists. (Distributing folios to other districts reduces the chance of panels being familiar with school programs.)

On the basis of the submitted folios, the panel indicates a level of achievement and a rung position for each sampled student, which can later be compared with the actual exit level (and the inferred rung placement) of these students. The panel also makes comments about the quality of the assessment program and the application of standards by the school. These comments for the 2001 random sampling exercise were part of the data used in this study to explore key terms associated with higher order thinking.

In addition to these core activities, the project also requires panellists to respond each year to a different 'special topic'. For the 2001 project, the special topic was higher order thinking skills. Random sampling panels were asked to answer three questions using a four-point scale.

Panels were also provided with space in which to make comments. These comments were also used in this project to identify key terms (see appendix three for the form used in the 2001 random sampling project).

FINDINGS

The readings of random sampling panels' comments made in 2001 on the basis of 2000 work, as well as the quantitative data gathered for the 2001 random sampling exercise, suggest what most teachers in Queensland know to be true. That is, that assessment of higher order thinking skills is one of the more challenging areas of assessment and that it is also an area that panellists more frequently refer to when they have advice to give a school about developing its assessment practices.

For the Board, and for schools, higher order thinking skills are significant because

- for equal opportunity in assessment, including assessment of higher order thinking, students need to have the best quality opportunities to show what they know and can do
- high-quality assessment packages, including sound opportunities to develop higher order thinking, are important to sound outcomes not just in terms of level of achievement decisions, but also Subject Achievement Indicators (for example, to make the finer distinctions between student work at the top end of the group, schools need the kind of evidence that sound opportunities to demonstrate higher order thinking can provide).

For the broader community, higher order thinking skills are important because of their relationship not simply to knowledge and skills required by further education and training and employment pathways, but also to their role in student self-development (for example, a key part of some literacy learning strategies is development of practices of self reflection, as documented in the 2000 Board report *Engaging Students in Literacy Learning*).

The quantitative and qualitative data can be described separately, although they are mutually reinforcing.

Quantitative data collected as part of the 2001 random sampling exercise

Responses to the questions asked in the 2001 Random Sampling Project (see appendix three for the data collection instrument) are summarised below.

Table one: Overall distribution of responses to items on higher order thinking

Does the school's approach to assessment	Not at all	Only a little	To some extent	A lot	No response
1. Encourage students to apply knowledge in demanding, unfamiliar situations (rather than rehearsed or routine situations)?	6 (1.6%)	75 (20.3%)	192 (51.9%)	86 (23.2%)	11 (3.0%)
2. Give students sound opportunities to complete complex, open ended, multifaceted tasks (rather than overly structured tasks with too many cues)?	16 (4.3%)	107 (28.9%)	164 (44.3%)	73 (19.7%)	10 (2.7%)
3. Allow students to be rewarded for demonstrating higher order thinking skills (such as thinking creatively, formulating and testing a hypothesis, using research material to analyse and synthesise, etc.)?	18 (4.9%)	99 (26.8%)	172 (46.5%)	69 (18.6%)	12 (3.2%)

How should we interpret the percentages given?

Some readers might point out that the findings are reasonably positive. After all, on each of the items the majority of submissions are represented by the panels as having an approach to assessment that provides 'some' to 'a lot' of scope of higher order thinking. Such a reader might point out that the fact that the bulk of submissions were found to provide 'some' opportunity for higher order thinking skills is not a cause for alarm because, after all, thinking skills are an important part, but not the major part, of what students do, and syllabuses do not suggest that higher order thinking skills are all that students should be required to do.

Yet such a reading is only part of the story of the quantitative data. It is likely the data are also saying that some schools are not providing students with the best possible opportunities to demonstrate higher order thinking skills. For each of the three questions about higher order thinking, roughly around a quarter of panellists indicated that the school's approach to assessment had little or no scope for higher order thinking skills.

There is another interesting question to ask. It is, 'What is the relationship between the random sampling data identifying serious problems with standards more generally, and the specific data about higher order thinking just given in table one?'

Table two gives the answer to this question. In the broader random sampling data there were sixteen submissions (out of 370) with serious difficulties with accountability or standards. The table suggests that while it is possible to have serious problems with standards or accountability in a submission that seems to offer students some or even a lot of opportunities for higher order thinking skills, most of the submissions with serious problems with standards tended to provide only a little or no opportunity for students to demonstrate higher order thinking skills. In other words, the table seems to reinforce the significance of assessment of higher order thinking in achieving comparability of standards.

Table two: Number of submissions with serious difficulties with accountability or standards, by panels' responses to questions about thinking skills

	Not at all	Only a little	To some extent	A lot	No response
Q1	1	8	6	1	0
Q2	5	3	6	1	1
Q3	4	5	6	1	0

Qualitative data collected as part of the 2000 and 2001 random sampling exercise

What were the nature of the comments about areas for improvement of higher order thinking skills made by district review panels as part of the 2001 random sampling exercise?

Such comments often seem to be about generic issues that cut across different subjects. The summary list below is not meant to imply any order of importance. The list suggests the range of different kinds of comments about higher order thinking skills; not all things in this list are equally applicable to all subjects, or all school contexts.

The list below captures the constructive criticisms made by panels; it does not reflect the many positive comments that panels made about submissions. In fact when we broadly categorise 2001 random sampling panels' comments on submissions made under the items about higher order thinking into positive, mixed, and only negative or constructively critical

comments we find that that there were eighty-three submissions that attracted only negative comments, 142 that attracted mixed comments, ninety-one that attracted only positive comments, and fifty-four submissions that attracted no comments at all. A submission that attracted only constructive criticisms/negative comments was not necessarily a submission that attracted the most negative rating on the quantitative scale discussed above.

Therefore, the list below represents all comments about areas of improvement noted by the panels under the items on higher order thinking, as well as any other relevant comments about the submission made elsewhere on the random sampling panel consensus form (whether or not the panel also made positive comments about other aspects of that school's assessment of higher order thinking or assessment package more generally, and regardless of the rating the panel made on the quantitative scale).

Panel comments about assessment of higher order thinking referred to

- *tasks not being sufficiently challenging, complex and multi-faceted* (i.e. complex tasks broken down into 'one step' tasks making them much easier)
- *assessment being 'routine' or 'practised'*, such that student responses are mechanistic and *'rehearsed'* in response to '*cues*' or *'hints'* in the item and do not use thinking skills
- *simply too much information* with an assessment task given in ways that prompt rather than extend students to demonstrate achievement at the top end
- assessment items being *too broad and lacking focus* so that students are not clear about what they are required to demonstrate
- *assessment tasks being too structured or 'multi-step' or 'segmented'* (i.e. too much 'direction' or 'prompting' or structuring of tasks or breaking complex tasks into smaller components) in ways that do not maximise opportunities for students to demonstrate self-directed higher order thinking
- assessment situations not providing students with opportunities to apply knowledge and skills in *unfamiliar or unseen or novel situations*
- assessment situations not being *authentic* enough or too 'unreal' for students to apply complex thinking skills
- *not enough information given to students* about the nature of what they are supposed to demonstrate in this area (i.e. through criteria sheets before a task, and through teacher feedback after the task is completed)
- *inappropriate classification of assessment items* (e.g. classifying items under criteria that are not about higher order thinking so that they are inappropriately used in determining overall achievement)
- simply not enough assessment of higher order thinking in the total assessment package
- lack of scope for students to respond 'creatively' or with 'originality'
- assessment being too knowledge or content based, i.e. not allowing scope for analysis, evaluation, and interpretation
- assessment being too focused on a set of 'core' knowledge and skills to allow scope for students to extend thinking skills
- assessment instruments not giving students enough scope to apply content knowledge to an unfamiliar problem
- assessment leaving out key topic areas in which thinking skills could have been applied
- assessment instruments leaving out some key aspects of higher order thinking skills
- assessment situations not providing students with enough scope to formulate and test a hypothesis
- practical or 'hands on' activities not giving students enough scope to demonstrate higher order thinking

- questions accompanying stimulus material/texts not taking full advantage of the scope for higher order thinking suggested by the stimulus material
- students not using research material/data/approaches to devise, plan, test, analyse, synthesise, and justify in self-directed research tasks, i.e. either not enough scope for research or else using research to reproduce content knowledge
- assessment instruments that require students to outline rather than use skills of analysis to compare and contrast
- instruments not giving students scope to express a particular opinion after having 'weighed' up the evidence
- lack of scope for *justification*
- *assessment mode* (multiple choice, short-answer questions, essays, etc.) not allowing students to demonstrate particular higher order thinking skills or else not enough different kinds of modes to provide students with suitable opportunities to demonstrate their strengths or problems (e.g. too many multiple-choice questions with only one 'right answer' so that, in the total assessment package, there is not enough opportunity for 'open ended' problem solving)
- *lack of variety of assessment within these modes* not allowing students to demonstrate particular higher order thinking skills
- *lack of variety of genres across assessment package* not allowing students to demonstrate particular higher order thinking skills
- lack of *alternative assessment items* giving sufficient scope for students to apply knowledge in different life-related situations
- lack of scope for students to respond in sufficient *detail and length*, e.g. lack of scope for extended written responses that demonstrate higher order thinking
- assessment being 'closed' rather than 'open ended', i.e. assessors having an expected 'right answer' but not providing scope for, acknowledging, different ways in which students can provide a right answer that demonstrates complex reasoning
- *interpretation of assessment criteria* related to higher order thinking skills not reflecting a sufficiently high standard, particularly at the top end of the student cohort
- *lack of clear understanding by assessor of the meaning of assessment criteria* relevant to higher order thinking skills leading to poor assessment tasks and/or inappropriate or inconsistent assessment judgments
- *assessment schema* not providing sufficient scope for more complex analyses in student responses
- the allocation of 'marks' to specific items (i.e. 'weighting' of 'marks') not suitably reflecting the importance of higher order thinking skills; this is related to assessors placing too high a value on 'repetitive mundane tasks' across an assessment package
- *design problems in assessment instruments* (such as excessive 'wordiness', ambiguity, 'obscure' situations) *and assessment conditions* (e.g. insufficient time given to complete response) presenting barriers to students' being able to demonstrate higher order thinking skills
- lack of opportunities for students to *use technology* to demonstrate reasoning skills in demanding and unfamiliar contexts
- assessment instruments that do not allow *less academically able students* to demonstrate higher order thinking
- lack of *appropriate constructive teacher feedback* in ways that help students extend their higher order thinking skills.
- Such comments often suggest that assessing higher order thinking skills has a particular role to play in *discriminating* between students, particularly at the top end (but also at the threshold of a particular level of achievement). Considered collectively, these

comments do not suggest there is always or nearly always a *single* feature or particular group of features of assessment (such as those relating to assessment instrument design) that is lacking when suitable opportunities for assessment of higher order thinking are not provided. The comments suggest that when different assessment items do not work successfully to provide opportunities for higher order thinking this can be the outcome of one or more aspects of the assessment approach.

WHAT MIGHT THESE TERMS AND CONCEPTS MEAN?

The following discussion makes very brief reference to some literature in the area of assessing higher order thinking, to help 'flesh out' the possible meanings of some key terms and concepts in assessing higher order thinking. It is organised under some of the words or concepts to do with higher order thinking that have been mentioned previously in our summary of the comments of the 2001 random sampling panels. It is not intended to be prescriptive and we have purposely restricted the discussion to very few writers on the subject and only some terms and concepts. This is because we wanted to convey the idea that this discussion is only a sample of the large amount of literature on this subject. We specifically want to avoid suggesting that there is a single 'endorsed' set of writers on this subject because teachers would be best served by reading widely to obtain a sense of many different approaches. Appendix two provides an annotated list of references that should be of interest to teachers who want to explore these concepts further; it is a big enough list but even so it is not exhaustive.Assessment that is challenging, complex and multi-faceted

What does challenging, complex and multi-faceted mean? Wiggins (1992) argues that

> … we would do well to use Lauren Resnick's criteria in our search for better-designed instruments. Higher order thinking
> - is *nonalgorithmic*—that is, the path of action is not fully specified in advance;
> - is *complex*, with the total path not visible from any single vantage point;
> - *often yields multiple solutions*, each with costs and benefits;
> - involves *nuanced judgment* and interpretation;
> - involves the *application of multiple criteria*, which sometimes conflict with each other;
> - often involves *uncertainty*, because not everything that bears on the task is known;
> - involves *self-regulation* of the thinking process, rather than coaching at each step;
> - involves *imposing meaning*, finding structure in apparent disorder;
> - is *effortful*, with considerable mental work involved.

Robert Glaser (1991) argues that the emphasis in assessing competence in subject matter should not be on 'the accumulation of facts and their reinforcement' but rather on 'the structure and coherence of knowledge and its accessibility in problem solving and reinforcement.' (p. 28). He argues that this 'competence in a subject matter' is displayed by 'evidence of a knowledge base that is increasingly *coherent, principled, useful* and *goal-oriented*' (p. 26). This means that if assessment items are going to capture this high-level competence, they need to focus not so much on 'isolated definition and superficial understandings' but rather target larger, structured interrelated 'chunks of knowledge', identification of 'principles that lie beneath apparent surface features', relation of information 'to the goals or problem solving and conditions for action', as well as 'self-regulatory skills' (p. 27). Glaser describes self-regulatory skills as skills for self-monitoring of performance, such as checking the appropriateness of 'problem-solving tactics', 'judging problem difficulty', time management, prediction of the outcomes of one's own performance and so on.

Designing tasks that allow students to analyse, synthesise, and justify

What does it mean to say a task gives students sufficient scope to analyse, synthesise and justify? Wiggins (1992) has this advice:

> Think of the knowledge to be tested as a tool for fashioning a performance or product. Successful task design requires making the essential material of a course a necessary means to a successful performance end … We want to know: Can the student use knowledge and resources effectively, *to achieve a desired effect*? (27)

Designing assessment situations

Wiggins (1992) argues that it is important to properly contextualise the assessment task.

> The aim is to invent an authentic simulation, and like all simulations, case studies, or experiential exercises the task must be rich in contextual detail. A context is rich if it supports multiple approaches, styles, and solutions and requires good judgments in achieving an effective result. One must please a real audience, make a design actually work, or achieve an aesthetic effect that causes pride or dismay in the result. (27)

In short, as Wiggins concludes here, the assessment task

> … may be a contrivance, but it needn't *feel* like one … A context is realistic to the extent that we so accept the premises, constraints, and 'feel' of the challenge that our desire to master it makes us lose sight of the extrinsic factors and motives at stake—namely that someone is evaluating us. In just this way, for example, putting out a school newspaper for a journalism course doesn't feel contrived. (27–28)

Assessment that is not about 'cueing' rehearsed responses

Wiggins (1992) has argued that student performance 'is not just doing simplistic tasks that cue us for the desired bit of knowledge. It entails "putting it all together" with good judgment; good judgment cannot be tested through isolated, pat drills.' (28) He argues that

> … we should consider the difference between drilled ability vs performance ability and ask: What is the equivalent of the game or recital in each subject matter? What does the 'doing' of mathematics, history, science, art, language use, and so forth, look and feel like in context? What are the projects and other kinds of synthesizing tasks performed all the time by professionals, consumers, or citizens that can be adapted to school use? Such tasks are always 'higher order' … (29)

Using different assessment modes

An over-reliance on traditional modes of assessment (such as tests) can work against assessment of higher order thinking. Wiggins (1992) has argued that

> Typical tests, even demanding ones, tend to overassess student 'knowledge' and underassess student 'know-how with knowledge'—this is, intellectual performance. Auditing local tests with Bloom's taxonomy as criteria, for example, shows that synthesis is infrequently assessed at present, and is inherently resistant to assessment by multiple-choice tests because it requires 'production of a unique communication' that bears the stamp of the student. (27)

Designing assessment conditions

The view in the literature about what is best practice in assessment conditions often relates to the authenticity of assessment conditions. The constraints typically relate to time, reference material, access to people, and access to information about the assessment situation. Wiggins (1992) points out that 'traditional testing, because it involves indirect proxies for performance, requires numerous inauthentic constraints to preserve validity.' (30) For Wiggins, finding sound constraints or conditions of assessment involves answering this question: 'What kinds of constraints authentically simulate or replicate the constraints and opportunities facing the performer in context?'

Providing information to students

Assessment of higher order thinking should involve sound information to students about the nature of what is being assessed, using explicit criteria. ‘A complex task is not a vague task’ Wiggins (1992, 29) has argued:

> All real-world performers know the target and the standards, not just their task in advance; such knowledge guides their training and rehearsals. Students should never have to wonder ‘Is this right?’ ‘Am I finished?’ ‘How am I doing?’ ‘Is this what you want?’ In a ‘real’ problem the task is ill-structured but well-defined: the goal, specifications, or desired effect is known, but it is not obvious how to meet it. (29)

Valuing higher order thinking in assessment decisions

Wiggins (1992) has argued that we should assess the substantive and important aspects of student performance, not ‘what is easiest’ to assess, i.e. what is easiest to observe (29–30). He adds that the finest assessment task can be rendered ‘bogus’ if we base assessment on the obvious and superficial. He argues that there are at least two things the assessor needs to identify: the most important features of a student response for each level of achievement, and the aspects of a student response that are most justifiable for use in lowering a student result (30).

WHAT SHOULD I DO ABOUT DEVELOPING ASSESSMENT OF HIGHER ORDER THINKING?

The Board has developed a workshop publication that schools can use to examine what Senior was like from the whole student or cross-curriculum perspective (*Reviewing Senior from a Student Perspective: A workshop for teachers*). Such a workshop has been used extensively by Board facilitators to support schools that want to develop practices, including the extent to which the school is providing opportunities for higher order thinking skills (this is not the sole focus of that workshop, but it has certainly been one of its outcomes for schools that scrutinise folios as part of this workshop.)

This document provides another workshop because, while the reviewing senior workshop is useful for getting a cross-curriculum perspective on practice, it does not provide the detailed examination of the design of particular assessment instruments that is also needed to develop practices in the area of assessing higher order thinking.

How should assessment instruments be scrutinised to develop the extent to which they assess higher order thinking skills? Clearly such a workshop cannot really have a narrow focus if it is to be useful—the discussion in previous sections of this report suggests how many aspects of assessment are important to achieving sound opportunities for assessment of higher order thinking skills. For a school to develop assessment of higher order thinking, more is required than, for example, simply having the list of common curriculum elements that relate to particular higher order thinking skills. We need to be able to look more deeply into practice, because all the checklists in the world cannot tell us *how* to improve the extent to which a particular instrument assesses higher order thinking skills.

Appendix one gives a workshop on assessment practices that has been trialled by Board researchers to help schools that want to use a whole school approach to developing assessment that has depth. It involves staff of a school bringing along one folio (with student work and assessment instruments) in a subject in which they have assessment expertise. From this folio staff select a substantial assessment task. They scrutinise the task and the student response to the task and fill in a tasksheet that helps them reflect on the match between intention and practice. They then work in a group to identify a checklist of things that could be done to achieve greater depth in the assessment items they develop, and a better match of intention and practice. At the end of the workshop one spokesperson from each group shares the findings of the group and the forms are collected and given to the school principal.

This instrument was trialled with seven schools and we found it often did focus on the kinds of issues about assessment of higher order thinking discussed in this report. Such a workshop can help a school pool assessment expertise through collaboration to develop practical suggestions for improving practices. The workshop focuses at the micro level on assessment instruments being used in the here and now at a school. Improvements to instruments scrutinised in the workshop and in use in the school are one immediate outcome of such a workshop. Another, less tangible outcome, is helping a school create a climate of critical self-scrutiny of assessment instruments which the national guidelines for assessment quality and equity (provided with the workshop materials) strongly emphasise.

Teachers will, of course, want to consult other resources for ideas about suitable workshops. Our experience in developing and trialling this workshop is that scrutinising assessment instruments for the match between intention and practice can be one of a number of ways of exploring how to achieve depth in assessment practices, including improved assessment of higher order thinking. Appendix one gives complete details of the workshop and supplies all the necessary materials so that schools can run this workshop themselves if they so choose.

REFERENCES

Allen, J. R & Bell E. J. 2000, *Reviewing Senior from a Student Perspective: A workshop for teachers*, Queensland Board of Senior Secondary School Studies, Brisbane.

Bell, E. J. & Allen, J. R. 1996, *Studying Assessment Practices: A resource for teachers in schools*, Queensland Board of Senior Secondary School Studies, Brisbane.

Engaging Students in Literacy Learning, 2000, Queensland Board of Senior Secondary School Studies, Brisbane.

Glaser, R. 1991, 'Expertise and Assessment', in Wittrock, M. C. & Baker E. L. (eds), *Testing and Cognition*, Prentice Hall, New Jersey, 1991.

Resnick, L. 1987, *Education and Learning to Think*, National Academy Press, Washington, D.C.

Wiggins, G. 1992, 'Creating tests worth taking', *Educational Leadership*, vol. 49, no. 8 (May), pp. 26–33.

APPENDIX ONE:
ASSESSMENT PRACTICES: A WORKSHOP FOR TEACHERS

Why have this workshop?

In this workshop we aim to:

- develop better teaching and learning for students through improving assessment practices in individual classrooms and across subject areas
- find workable ways to get more clarity, depth and equity in assessment practices.

What do we need to run this workshop?

Requirements for this workshop are:

- a facilitator
- an allocated time of about two hours
- participants—teachers who have each nominated a teaching subject for particular focus, i.e. the workshop participants can be from any different subject areas in a school
- for each participant, one complete student folio of assessment task responses, preferably for a student who has achieved well in the subject nominated, and all folios coming from the same student cohort
- tables and chairs arranged for groups of four participants
- a copy of this workshop paper for each participant.

What happens in the workshop?

1. The facilitator provides a brief introduction **(15 minutes)**:
    - explaining the assumptions on which the workshop is based
    - explaining the concepts of clarity, depth and equity in assessment practice.
2. Teachers select and revise an assessment task **(55 minutes)**.
3. Teachers develop key rules/checklist questions/processes they can use to improve assessment practice (**30 minutes)**.
4. Teachers discuss their findings with the whole group **(20 minutes)**.

What are the key assumptions?

We assume that:

- thinking and reflecting are key ingredients in effective practice
- teachers need workable strategies that are quick and efficient
- good assessment tasks are developed through an iterative process in which teachers:
- select an initial idea/concept/starting point
- critique and review
- revise and polish
- repeat the last two steps as required
- practice sends more powerful messages than do statements of intention
- an assessment task begins with its development by the teacher, continues with its interpretation by the student and ends with the teacher's judgment and feedback to the student.

What do we know about good assessment practices?

The facilitator provides input on assessment practice ***(15 minutes)***.

We know some things about good practice in assessment:

- the purpose of an assessment task is to give each student an opportunity to show what the student knows and can do
- assessment tasks should ask the student to demonstrate knowledge, understanding and skills directly related to the relevant elements of the curriculum
- we should tell students about the criteria and standards by which achievement will be judged before they do the task
- 'clarity' means that what you have to do is explicit, the task uses clear, direct language and the layout helps access
- an assessment task has 'depth' when it requires the student to use processes such as hypothesising, criticising, analysing, justifying
- if the task is set in a 'real-world' context, the context should be an integral part of the task, not 'window-dressing' or an obstacle to accessing the task
- we can use evidence, reading back from an assessment item to the in-practice values, to examine an assessment task, i.e. look at the nature of the good response to see
- the depth required (in practice)
- the quality required (in practice)
- the quantity required (in practice)
- the match between what the student did and the clues/cues etc. in the item or inferences in the general background.

The ACACA principles for quality and equity in assessment provide useful statements of good practice (these guidelines are included in these workshop materials).

What do teachers do in the workshop?

The facilitator issues the workshop paper and explains the purpose of this activity ***(5 minutes).***

Examine assessment practice in your subject in order to identify:

- how you might improve assessment practice
- how you can improve teaching practice through improvement in assessment practice.

The facilitator explains the process of the first task ***(5 minutes).***

Using Tasksheet 1, review a particular assessment task in your subject to identify:

- what knowledge, understandings and skills were assessed
- what was valued in the assessment
- elements of good assessment practice
- some processes that you could use to enhance assessment practice and teaching practice.

The facilitator sets out the procedure for this activity ***(5 minutes).***

In the group of four, take about **10 minutes** to look at the folios to select an assessment task for further analysis, i.e. a substantial task, one that has some intended depth, one that students did well on, with some marking by the teacher.

Now take about **30 minutes** to fill in **Tasksheet 1**, looking at the task in terms of:

- what the student actually had to do to get rewarded, i.e. look at what gained ticks or comments of approval

- what was most or least valued in the item, i.e. what gave a student best or worst return for effort
- what were the assumed and background knowledge and skills required for the task
- how the student was informed of what was required, i.e. how did the student know the nature of the good response, the depth wanted, the quality, the quantity? (Find the match between what the student did and the clues or cues in the item or inferred from the general background.)
- how the assessment task could be revised to achieve more clarity, depth and equity.

*The facilitator sets out the procedure for the next stage of the workshop **(5 minutes)**:*

In the group of four, take about **30 minutes** to fill in **Tasksheet 2**, devising about five key rules/checklist questions/processes you could use on an everyday basis to:

- enhance the match of intention and practice in assessment tasks
- enhance the depth of learning required by assessment tasks
- enhance students' understanding of the assessment tasks you want them to do.

Nominate a spokesperson for your group. Then, with the whole group of participants, take about **20 minutes** to share, discuss and summarise your findings and to ask any questions you have for each other or for the facilitator.

What next?

The facilitator, senior administrators or heads of department collect the summaries and other materials developed to reflect on later and use in developing good practice further.

TASKSHEET 1 — REVIEWING A TASK

Look at the folio to select an assessment task for further analysis, i.e. a substantial task, one that has some intended depth, one that the student did well on, with some marking by the teacher. As you read these, comment on the task in response to the following questions:

What did the student actually have to do to get rewarded, i.e. look at what gained ticks or comments of approval?

What was most/least valued in the item, i.e. what gave students best/least return for effort?

What were the assumed and background knowledge and skills required for the task?

Tasksheet 1

How was the student informed of what was required, i.e. how did the student know/work out/infer the nature of a good response, the depth wanted, the quality, the quantity? (Find the match between what the student did and the clues or cues in the item or inferred from the general background.)

How might the assessment task be revised to achieve more clarity, depth and equity?

TASKSHEET 2 — IMPROVING ASSESSMENT PRACTICE

Devise about five key rules/checklist questions/processes you could use on an everyday basis to:

- enhance the match of intention and practice in assessment tasks
- enhance the depth of learning required by assessment tasks
- enhance students' understanding of what it is you want them to do.

Workshop handout: Some suggestions when designing an assessment task (use this as background reading to help 'brainstorm' ideas about assessment)

- Identify the curriculum elements (knowledge, skills, understanding etc.) that the task will assess.
- If a 'real' context is used ensure that it is not just 'window dressing' and that it does not present a barrier to access to the task.
- Give clear and definite instructions about what students are to do.
- Give clear and specific requirements for the task (mode of response, length, duration etc.).
- Include cues to resolve uncertainties about what the task requires.
- Specify the conditions in which the task will be undertaken.
- Consider using some of the general terms used in the QCS Test to familiarise students with aspects of that test.
- Devise a format and layout for the task that helps the students to see what the task expects of them in terms of demonstrating what they know and can do.
- Ask other teachers to review the task for clarity, depth and equity.
- Ensure that students will know beforehand what this assessment item assesses.

Workshop handout: Some suggestions for use when developing task-specific standards (use this as background reading only)

- Identify the syllabus criteria that will be assessed by the task.
- Describe the nature(s) of fully successful responses.
- Describe the natures of partially successful responses.
- Specify for each standard what students will be *doing* to *what* in what kind of *situations*, i.e. verbs (or processes), nouns (or participants) and circumstances, ensuring that these are specific to the task.
- Describe the standards so that they correspond to real differences in achievement that are readily discernible in student work.
- Describe the standards so that they allow teachers to make sound judgments about distinct features of student responses at the different levels.
- Use clear, direct language that students will readily understand.
- Identify and resolve any ambiguities in using and interpreting the descriptions of standards.
- Develop a clear interpretation of the terms used in the description for each standard.
- Provide enough detail in the descriptions of standards to ensure that students will not be inappropriately penalised.
- Consider the range of different student responses that could be awarded the same grade.
- Consider the different forms of expression that student responses might take.
- Consider developing an exemplar (that is, an example of an ‘A’ student response).
- Consult other teachers to see that the standards are reliable (different teachers marking the same student work will award the same grade).
- Identify and set aside any issues such as:
- personal expectations of individual students
- preconceived notions of the distribution of grades
- issues that are not relevant to what the assessment item is assessing.
- Ensure that students have access to the overall criteria and standards you will use to judge their work before they undertake the task.

A C A C A
Australasian Curriculum, Assessment and Certification Authorities

Guidelines for Assessment Quality and Equity

Context and scope of these guidelines

ACACA agencies are responsible, directly or indirectly, for enormous amounts of assessment—hundreds of examinations, thousands of questions, tens of thousands of assessment items set in schools—in the high stakes area of senior secondary assessment.

Senior secondary assessment has a significant impact on the lives of the students involved. Accordingly, they each deserve a fair go—the same opportunity to show what they know and can do, regardless of factors such as gender. Fairness—equity—and quality are closely related, not least in the sense that reasonable quality is a prerequisite for the deliberate achievement of equity. Low-quality assessment will almost certainly lead to inequitable outcomes.

There is, of course, always room for improvement in the quality of assessment. Each year sees an increase in the sophistication of educational assessment.

The adoption and implementation of these guidelines represent a step forward in improving the quality, and hence the fairness, of assessment. They do not represent a definitive, final or complete position. They should be read as representing a commitment by ACACA agencies to moving forward by setting down some principles and adopting nationwide an agreed position. The guidelines are intended to be a starting point and so should be read as being deliberately both provisional and incomplete. Experience with, and evaluation of, the effect of their adoption will lead to their further development and enhancement.

ACACA agencies are responsible for assessment, not for how schools operate. The need to be fair to all students means that ACACA assessment practices should not seek to adjust for missed opportunities to learn by saying that students have demonstrated achievement that they have not in fact demonstrated. In this sense the equity issues connected with ideas of Opportunity to Learn are not, therefore, part of these guidelines.

These guidelines are concerned with the practice of assessment. Syllabuses specify the knowledge and skills that should be assessed. The task of assessment is to assess in terms of these specifications, not to redefine the syllabus. However, the guidelines include the need for systematic and regular evaluation and review of assessment: its methods, its materials and its results. The systematic evaluation and review of assessment in a subject provides information about the syllabus: information that should be fed back into the syllabus review and development process.

Fundamental to equity in assessment is the recognition that the construction of the knowledge and skills to be assessed should involve a critical evaluation of the extent to which the choice of a particular set of knowledge and skills is likely to privilege certain groups of students and exclude others by virtue of gender, socioeconomic, cultural or linguistic background. A concern with equity also leads to adopting a proactive stance on the appropriate representation in the curriculum of different kinds of cultural knowledge and experience as valued knowledge and skills.

Guidelines for assessment quality and equity

(Note: **Bold** text indicates references to the guidelines.)

These guidelines are about the quality of assessment - methods, materials and results - and hence implicitly and explicitly about equity. The principal touchstone of quality in assessment is the extent to which it gives students a fair go .

To **certify**[1] achievement in a subject requires **assessment** of **students**' command of the **knowledge and skills** defined and required by the syllabus. This assessment occurs through **a set** of assessment instruments. **Assessment instruments** include such devices as supervised examinations, assignments, projects, practicals, orals, aurals, observational schedules and portfolios. Assessment occurs under various **conditions** - supervision, notice, access to resources, times, dates, handing-in procedures and acceptance of late submissions. Each assessment instrument comprises one or more **assessment items**. An assessment item requires a response from students in one or more of a variety of **modes** such as multiple choice, short answer, paragraph, extended written response, oral, graphical, diagrammatic and so on. As well as the knowledge and skills that are the direct focus, assessment items draw on **assumed knowledge** and often make use of **background material** contextualising the task presented. **Cues** and the **layout** of assessment items provide guides to students about the requirements of the task set by the item. Assessment items are devised by **item writers** and selected and combined into assessment instruments by **test designers**[2].

1 Agencies may need to replace some or all of the highlighted terms with the terms they currently use.
2 Frequently, of course, the test designer is also the item writer.

These guidelines were endorsed at the Australasian Curriculum, Assessment and Certification Authorities (ACACA) meeting in August 1995.

The development of the guidelines was funded by the Commonwealth Department of Employment, Education and Training.

Printed by the Queensland Board of Senior Secondary School Studies,
PO Box 307, Spring Hill, Qld 4004, Australia.

**For more information, contact Reg Allen,
Deputy Director—Analysis & Resources, QBSSSS,
telephone (07) 3864 0259, fax (07) 3221 2553.**

RE0654

For quality, and hence equity—

Each assessment item should:

- actually assess what the item writer intends it to assess
- assess what on face value it purports to assess
- only require for its successful completion the decoding of a subtext after critical examination of the necessity and appropriateness of this requirement
- use specialist language or jargon as an aid to clarity and accuracy and not as an irrelevant obstacle to successful demonstration of the required knowledge or skill
- only involve the reproduction of gender, socioeconomic, ethnic or other cultural stereotypes after careful consideration of the necessity of such reproduction
- in order to avoid being itself a barrier to students' demonstrating their command of the characteristic the item is supposed to assess:
  - be clearly presented through appropriate choice of layout, cues, visual design, format and choice of words
  - state its requirements explicitly and directly
  - use as background material and require as assumed knowledge only that to which students may reasonably be presumed to have had ready access
- be marked by explicit, clear, unambiguous, criteria declared in advance that:
  - allow the student to identify appropriate ways to demonstrate command of the required knowledge and skills
  - allow the marker to recognise, where appropriate, different ways in which a student may demonstrate command of the required knowledge and skills

Each assessment instrument should:

- give students clear and definite instructions
- be used under clear, definite and specified conditions that are substantially the same for all
- be used under conditions that do not present an inappropriate barrier to the equal participation of all

Each set of assessment instruments used to assess a student's achievement in a subject should:

- involve the use of a range and balance of background contexts in which assessment items are presented
- involve a range and balance of types of assessment instruments and modes of response, including a balance and range of visual and linguistic material

  involve a range and balance of conditions

Each agency should endeavour to improve assessment practices for which it bears responsibility:

- by regular review of the methods, materials and results of assessment
- by including an examination of equity in quality control and quality assurance procedures such as research and data analysis
- by training item writers, test designers and markers directly employed by the agency
- by communicating its quality requirements to suppliers of assessment items and instruments
- by promoting the adoption of these guidelines by teachers responsible for school-based assessment contributing to results certified by the agency.

An illustrative list of strategies for developing improved practices in assessment emphasising quality, equity and gender equity

Agencies can:

- distribute summaries of relevant research findings to item writers and test designers
- conduct reviews of the representation of gender in assessment instruments
- evaluate the occurrence in assessment instruments of reproductions of gender, socio-economic, ethnic or other cultural stereotypes
- evaluate the consistency of the syllabus and the total set of assessment instruments by identifying the relative importance given by each to particular knowledge and skills
- review the extent to which the set of assessment instruments in a subject are, to the full extent permitted by the syllabus, inclusive of the experience and achievements of women as well as of men
- provide item writers and test designers with examples showing the full extent to which the construction of knowledge and skills to be assessed allows assessment incorporating references to human contexts
- conduct equity scanning of assessment instruments before use
- promote research into the validity and fairness of assessment items for which the agency is responsible
- scrutinise assessment instruments to eliminate material that may prove an unnecessary and irrelevant distracter for some students
- employ specialist editors to examine the language of assessment instruments in terms of possible barriers to equal opportunity for all students
- involve students in critical examination of the quality of assessment instruments they have experienced
- involve teachers in critical examination of the quality of assessment instruments their students experience
- involve parents and the wider community in reviews of assessment practices
- review marker recruitment, selection and training procedures
- evaluate the weighting of assessment items and instruments in terms of analyses of gender differences in student performances.

APPENDIX TWO:
ANNOTATED LIST OF RESOURCES

Annis, D. & Annis, L. 1979, 'Does philosophy improve critical thinking?', *Teaching Philosophy*, 3:2 Fall, 145–152.

Annis and Annis discuss the proposal that studying philosophy as a subject will improve a student's ability to think critically. This could lead to an investigation of the instructional factors involved in teaching and learning in this discipline, which could be used in general courses on critical thinking.

Arter, J. A. & Salmon, J. R. 1987, *Assessing Higher Order Thinking Skills: A consumer's guide*, Northwest Regional Educational Laboratory, Portland, Oregon.

This handbook examines higher order thinking skills (HOTS) in terms of how these can be identified, and reviews some testing items used to assess these. There is a useful matrix of thinking skills and a checklist for selecting appropriate test items. The test items reviewed are not particularly relevant to our situation.

Benderson, A. 1990, 'Critical Thinking: Critical issues', *Focus: Educational Testing Service.*

The introduction to this text explores the tension between the critical thinking theories developed by philosophers and by psychologists. The kinds of problem solving developed by philosophers seem more in tune with those used in education because unlike the logic and number puzzles developed by psychologists, which have a correct answer monological, their problems require a wider range of thinking skills multilogical. One way to resolve the tension, proposes Benderson, is to increase student involvement in learning—to make it more active. He also examines initiatives in various US states to make critical thinking part of the curriculum and to test these skills.

Berlak, H. 1965, 'The teaching of thinking', *The School Review*, 731, Spring.

An early article which summarises much of Dewey and Ennis's work. Berlak presents a case for avoiding the teaching of a generalised structure of thinking processes on the grounds that thinking processes differ according to the demands of the subject matter or domain and that thinking processes are as yet not completely understood. If educators concentrate on context-specific models, they may be able to develop pedagogies and material that will allow them to teach critical thinking/reflective thinking/problem solving.

Beyer, B. K. 1984, 'Improving thinking skills—Practical approaches', *Phi Delta Kappan*, April, 556–560.

A precursor to the 1985 article. A strong argument is made for the sequential teaching of critical thinking as part of the subject-based curricula. Beyer believes that if we can identify the steps involved in a thinking skill then it can be taught effectively. This would occur in four stages: introduction, reinforcement, extension and practice.

Beyer, B. K. 1985, **'**Critical Thinking: What is it?', *Social Education*, April.

In this article Beyer focuses on arriving at a definition of thinking skills, arguing that unless they are defined, they cannot be taught effectively. He examines the theories proposed by a range of critical theorists and either dismisses them as too vague or incorporates their arguments into his own definition and ideas for application. He identifies ten essential skills or operations and maintains that these are the core of critical thinking.

Beyer, B. K. 1987, *Practical Strategies for the Teaching of Thinking*, Allyn & Bacon, Boston.

In chapter 9 in this text, 'Assessing Student Thinking', Beyer argues that for students to value the higher order thinking skills being taught, they must be assessed. He believes there are two ways that this can be done so that the quality of thinking is assessed, and changes in the quality are evaluated. The first is through in-class testing that moves beyond recall of subject

matter and the second is through classroom observation. There is an example of the kind of in-class testing where items testing a new skill are grouped. Several points arise in a discussion of these questions. These relate to the unfamiliarity of the data, how much needs to be processed and the emphasis on the central concern with how students use the data to inform their thinking processes. Beyer also emphasises that the skills cannot be transferred to new contexts until students have had considerable practice in the skill. In terms of observing students' development of critical thinking, he cautions that this form of evaluation is meant only to support other judgments and should never be transformed into comparative assessments of student skills.

Beyer, B. K. 1997, *Improving Student Thinking*, Allyn & Bacon, Boston.

This text is readable and interesting and focuses on classroom practice. Beyer establishes his central argument that four elements must be built into curriculum:

- Establish and maintain a thoughtful classroom—one that nourishes thinking as well as minimises the risks inherent in efforts to engage in and improve one's thinking.
- Make visible and explicit the thinking of our students and of others already skilled in doing it.
- Guide and support student efforts at thinking by various techniques that serve to provide continuing scaffolding and cues for that thinking.
- Integrate the practice and use of and instruction in thinking with meaningful instruction in major subjects across the curriculum.

The text then explores each of these elements and how each contributes to the development of student thinking at the higher order levels. The suggestions and models are practical and student focused and support Grant, McPeck and Siegel.

Bransford, J. D., Burns, M. S., Delclos, V. R. & Vye, N. J. 1986, **'**Teaching Thinking: Evaluating evaluations and broadening the data base', *Educational Leadership* 44, October, 68–70.

This article examines the evaluation of thinking programs. Transfer is an important concept and it is stressed that this cannot be measured without first assessing whether the skills have been learned in the initial program.

Burkhalter, N. 1993, 'How persuasive writing aids critical thinking', *Speech Communication Association*, November 18–21.

Burkhalter explores the connection between writing skills and developing critical thinking skills. In particular, she examines the specific demands of persuasive writing which, due to the nature of its structure, links strongly to higher order thinking: so much so that she labels them as a subset. This argument is based on the process of persuasive writing that requires the writer to analyse, evaluate and synthesise arguments before s/he can produce the written product. In that process the writer further explores the ideas and develops new knowledge and understanding.

Cederblom, J. & Paulsen, D. W. 1996, *Critical Reasoning*, Wadsworth, Belmont.

The focus of this text is critical reasoning, which involves more specific processes. There is a section on the language used in identifying argument parts and a glossary of terms.

Costa, A. L. 1984, 'Thinking: How do we know students are getting better at it?', *Roeper Review*, 64.

Costa presents ten suggested characteristics which, when students are observed and anecdotal records are kept, may indicate or chart the intellectual growth associated with the development of critical thinking. Ideas can be connected to other writers through his

discussion of flexibility, meta-cognition, transference and the critical attitude. The emphasis on student behaviour makes this different in its approach from other theorists.

De Bono, E. 1982, *de Bono's Thinking Course*, BBC, London.

De Bono's interest is in developing thinking skills, and in particular on dealing with novelty. His program would be classified by Ennis as a general model in which thinking skills are taught separately from subject matter in disciplines. For each alternative solution to problems, de Bono suggests that one should list the positive *p*lus, negative *m*inus and *i*nteresting features of that solution. De Bono's emphasis is on lateral thinking as well as other more familiar aspects of the field.

Ennis, R. H. 1962, 'A Concept of Critical Thinking', *Harvard Educational Review*, 32, 81–111.

A paper in which Ennis proposes his base definition of critical thinking as 'the correct assessing of statements'. He then identifies twelve aspects of critical thinking. He does exclude creative thinking from this early definition and discusses three dimensions: logical, critical and pragmatic.

Ennis, R. H. 1981, 'Logic and critical thinking', *Educational Philosophy*.

This is a short response to McPeck's argument regarding the teaching of critical thinking skills as part of subject disciplines. Ennis agrees that subject-specific knowledge is indeed an essential condition for critical thinking but also questions whether it is **sufficient**. This is especially important in terms of the transfer of good thinking skills outside the subject area.

Ennis, R. H. 1989, **'**Critical Thinking and Subject Specificity: Clarification and needed research', *Educational Researcher*, 18(3) 4–10.

This article presents an overview of some critical thinking approaches: general, infusion, immersion and the mixed model. Ennis also examines three aspects of subject matter: domain, field, and idea, and comments on their meaning while discussing the problems created by vague understandings and use of these terms in discussions about critical thinking. With the concept of domain, the belief among theorists is that some background knowledge is needed to be able to think in a given subject area domain. There is also debate about whether thinking skills are transferred to other subjects or to everyday life if the ability to transfer is not taught explicitly and practised. In terms of field, a basic concept is that from field to field, definitions of what constitutes good thinking vary, and that the skills for each field should be central to its complete understanding. The third aspect of subject matter, the conceptual, is associated with McPeck and the immersion approach to critical thinking. This understanding is largely dismissed by Ennis.

Feely, T. 1976, 'Critical Thinking: Towards definitions, paradigms and research agenda', *Theory and Research in Social Education*, IV(1) August.

This paper focuses on the contribution of critical thinking to social studies. Feely establishes a definition that relates to the particular subject area based in part on Ennis's work and argues that whether or not critical thinking is taking place can be determined through the use of pre-designed criteria. He explores what he determines are the two paradigms of critical thinking: the logical and the mental and the implications these have for research into critical thinking and for the way it is taught.

Glaser, R. 1984, 'Education and Thinking', *American Psychologist* 392, 93–104.

The first part of this paper explores past theories as they relate to the development of thinking about higher order cognition. Various thinkers are examined and their ideas are related to current directions. Remembering the date of the paper, Glaser argues that thinking skills

programs which are unrelated to content/domain-specific knowledge tend to lack focus and applicability.

Grant, G. E. 1988, *Teaching Critical Thinking*, Praeger, New York.

This reference explores classroom practice in teaching critical thinking. The theoretical underpinning of critical thinking is explained in accessible language but the emphasis is on the connection to the work of teachers in classrooms. There is some common ground with McPeck in that Grant argues that thinking skills must be integral to classroom teaching and subject knowledge. The transformation of content into critical thinking tasks is discussed with regard to four case studies, presenting the methods used by teachers from different subject areas. This step is presented as an essential element of the progression from learning information to being able to use knowledge. A further assumption is that teaching critical thinking is a cognitive act that depends on three elements: pedagogical content knowledge, knowledge of students, and knowledge of self. For teachers who want to examine the way that some teachers are empowering their students by encouraging a thinking classroom, the chapter that examines the methods and style would be most helpful. This reference is quite dated but it is interesting and readable.

Haladyna, T. M. 1997, *Writing Test Instruments to Evaluate Higher Order Thinking*, Allyn & Bacon, Boston.

This text's title sums up its contents and is very specific to the techniques involved in test design. Haladyna explores the definitions he feels are essential to understanding student performance and there are explanations of a range of terms and instrument types. The ones most relevant to higher order seem to be the high inference items that require students to construct rather than select a response. There is an emphasis on the need to specify carefully what students need to learn and to design tasks or items that assess these demands. This text is a useful source for definitions and item explanations.

Halpern, D. F. 1996, 3rd edn, *Thought and Knowledge: An introduction to critical thinking*, Lawrence Erlbaum Associates, Mahwah, New Jersey.

The opening chapters of this text contain thoughts on the nature of knowledge and the value of critical thinking. The concept that in a time of rapid change we need to equip students with the capacity to cope with the 'deluge' of information is central. Also, Halpern examines the connection between knowledge and thought and how we can teach students to use the information they receive in daily life. One argument is that in the first two decades of their lives, our role as educators is to provide them with the basic skills and understandings to see them through the remaining 50+ years of their lives and that teaching them to think critically is essential. Critical thinking is seen as a tool which is a purposeful, reasoned and goal-directed cognitive process, one which has an evaluative function. The remainder of the text concentrates on the particular skills involved in critical thinking, information that is too specific for the purpose of this investigation. Her views about thinking confirm those expressed by other writers like Grant and Haladyna.

Hawes, K. 1990, **'**Understanding Critical Thinking', *Varieties of Thinking: Essays from Harvard's Philosophy of Education Research Centre*, edited by Howard, V. A., Routledge, New York.

Hawes seeks to test some of the possibilities for general application of critical thinking. There are summaries of the major writers/thinkers in the field: Dewey, Ennis, Paul, and McPeck. He specifies four aspects central to critical thinking: purpose, method, the thing evaluated, and the result of thinking. The discussion of method explores the idea of the learner developing a method that works for her/him and that method and purpose are connected. Part of developing as critical thinkers is being able to do this. This then connects to the other aspects, and he also discusses critical attitude and perspective.

Kennedy, M., Fisher, M. B. & Ennis, R. H. 1990, 'Critical Thinking: Literature review and needed research', *Dimensions of Thinking and Cognitive Instruction*, edited by Jones, B. F. & Idol, L., Lawrence Erlbaum, Hillsdale, New Jersey.

This is an overview of critical thinking as a theory and as an element of educational instruction. Although Ennis's input is clear, the coverage of a range of ideas and programs is comprehensive.

Laver, R. M. 1999, 'A Crisis for Educators: An opportunity for service', *ETC: A Review of General Semantics*, Fall 5632–52.

New standards introduced in New York demand that student knowledge moves beyond facts and into ways of thinking and dealing with facts. Laver examines the contribution that relational and critical thinkers can make in helping students and teachers meet these demands.

Lipman, M. 1988, 'Critical thinking—What can it be?', *Educational Leadership*, September 38–43.

This article focuses on the importance of criteria in learning and assessing critical thinking. One facet of the argument here is that discussions of critical thinking have so far failed to deal with the essential characteristics of what Lipman calls 'good' thinking. He explores these in more depth. There is extensive discussion of critical thinking in that it is self-correcting, sensitive to context, and leads to intellectual empowerment. In developing 'good thinking' he believes that all participants must learn to identify and cite good reasons for their opinions.

Marzano, R. and Pickering, D. 1997. Dimensions of Learning Teacher's Manual. Association for Supervision and Curriculum Development, Virginia USA.

This reference came highly recommended by a senior educator. It deals with developing higher order thinking skills across the curriculum and has been used as a reference in workshops for teachers, so there is plenty of material in it with a 'hands on' relevance. This publication develops the research and theory in a 1988 publication by Marzana et al (called Dimensions of thinking) into a practical framework for K–12 across different content areas. For this reason the manual should be of use to schools wanting to develop an across the curriculum approach to developing higher order thinking skills in teaching, learning, and assessment.

McPeck, J. E. 1990, *Teaching Critical Thinking: Dialogue and Dialectic.* Routledge, New York.

McPeck's views seem very pragmatic and straightforward and an unusual feature of this text is that he includes a critique of his ideas in the text, to which he then responds. The thrust of his arguments is that critical thinking is taught best through the subject disciplines and not as a separate 'thinking skills' course. The basis of this argument is that thinking can never be content-free. He also stresses the crucial role of language in developing autonomous thinkers, in terms of understanding the arguments presented and then evaluating them. He makes the strong statement that 'when disciplines are effectively taught, they provide the most fundamental and inescapable cognitive requirements for being rational' (41). However, the critical thinking will emerge in students' responses only when teachers ask for them and assess them. Like Beyer, McPeck believes that teachers must promote the value of the skills and open themselves to critical evaluation, in much the same way that texts are evaluated and questioned by students who are becoming critical thinkers, or as McPeck refers to them—'autonomous thinkers'.

Nickerson, R. S., Perkins, D. N. & Smith, E. E. 1985, *The Teaching of Thinking*, Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Section 11 focuses on the evaluation of programs that teach thinking and the processes involved. The next section examines what may be achieved if thinking skills are taught, starting form the basic premise that they are worth teaching. A number of considerations are discussed and recommendations are made.

Norris, S. P. 1985, 'Synthesis of research on critical thinking', *Educational Leadership* 42, May 1985, 40–45.

The title sums up the purpose of this paper. Norris examines a range of critical thinking theories, debates and tests. On p. 27 is a table listing highlights from research into critical thinking which is a useful summary.

Norris S. P. & Ennis, R. H. 1989, *Evaluating Critical Thinking*, Pacific Critical Thinking Press and Software, Pacific Grove.

Chapter 2 examines different techniques that enable assessors to gather information on students' critical thinking. The constructed response is favoured as a means of gathering more valid data and also because it allows students to demonstrate whether they can coordinate several thinking strategies to consider a complex problem. Observation, interviews and journals are other avenues explored. Techniques examined in more detail are: aspect-specific, comprehensive, general knowledge, subject-specific, and techniques with a variety of tasks Portfolio. There is emphasis on the critical disposition. The concepts of reliability and validity in testing are discussed in relation to gathering data on students' critical thinking. Subsequent sections deal with constructing multiple-choice test items, open-ended techniques and making decisions.

Pawlowski, D. R. 1997, *Challenging Students to Think: Making critical thinking and writing central to the basic course,* 84th National Communication Association, November.

This paper's concern is with the perceived need to make critical thinking speaking and writing central to college students' learning through a basic course in these skills at the start of their studies. There is some discussion of the history of critical thinking and the related theorists and there is a further connection to the communication processes that allow students to write or speak about their thinking. In other words, without expression, the thinking has no relevance and cannot be assessed.

Quellmalz, E. S. 1985, 'Needed: Better methods for testing higher-order thinking skills', *Educational Leadership* 43, October 29–35.

The author makes several recommendations about the way that higher order thinking skills should be taught and assessed if the information these processes use can be seen as valid. Some ideas: important issues in the domain; multiple interpretations and solutions to problems; formats that require an explanation of reasoning; assessing metacognitive skills. There is an important link made between instruction and assessment in terms of these skills. The figures used present useful information.

Resnick, L. 1987, *Education and Learning to Think*, National Academy Press, Washington.

In the introduction and first chapter, Resnick presents a historical perspective on higher order thinking skills. There are clear connections between ideas presented here and in later texts by Beyer, McPeck and Siegel. These can be made in the methods of teaching advocated and in the development of the critical disposition. This last element must, according to this author, be cultivated and students must receive opportunities to exercise it and support when taking the risks associated with higher order thinking. Another recommendation is that assessment should reflect the integrated thinking required when students are using higher order reasoning processes.

Siegel, H. 1980, 'Critical Thinking as an Educational Ideal', *The Educational Forum*, November 7–23,

Siegel is examining the position of critical thinking in educational theory, and in nominating it as an ideal is questioning whether it can be realised. Siegel, like McPeck, stresses the importance of the 'critical attitude', which they describe as a willingness or disposition to be critical. Impartiality of judgment and recognition of the force of reasons are central to critical thinking and for Siegel they also have a strong connection to moral education, which he sees as an integral purpose of education generally. He not only identifies and explicates his theory of critical thinking, but also justifies its use through the methods. The discussion also touches on the political nature of critical thinking but dismisses objections on the grounds that the liberating benefits of teaching critical thinking are too important. This article is a theoretical discussion so does not provide ideas for practical implementation or a method of teaching.

Sparapani, E. F. 2000, 'The effect of teaching for higher level thinking: An analysis of teacher reactions', *Education,* Fall 2000, 1211–80.

This article examines a research project in which teachers planned and delivered a series of lessons which demanded higher order thinking skills and then assessed the impact and results of these methods and activities. Changes reported included the impact on teachers in terms of the time needed to plan and the move to becoming mentors and facilitators, rather than dispensers of knowledge. Students appeared to have enjoyed the activities and moved beyond the lesson to make connections with other elements of the curriculum. They also searched for extra resources for themselves. System limits are also examined at a basic level.

Sternberg, R. 1985, 'Critical thinking: Its nature, measurement and improvement', *Essays on the Intellect*, edited by Link, F. R., ASCD, Alexandria.

This is an overview of critical thinking. The combination of three strands of theories, the philosophical, psychological and educational is discussed. Ennis, Lipman and Paul are linked to the philosophical strand which concentrates on the requirements of formal logical systems. Sternberg argues that this may not always be consistent with the capabilities of students in classrooms. These are better used as models of competence rather than performance in human thought. Psychologists are interested in the nature of human thought (Bransford, Bruner, Sternberg). Educationists like Bloom and Perkins focus on the skills that children need in classrooms and are closely tied to observation and experience.

Swartz, R. J. & Perkins, D. N. 1990, *Teaching Thinking: Issues and approaches*, Critical Thinking and Software, Pacific Grove.

This chapter examines the need to evaluate testing instruments which assess critical thinking skills. The emphasis is on evaluating programs in terms of design and formative and summative evaluation. There are several points listed for consideration in the design evaluation: skills dealt with; attitudes and styles; perception of need and opportunity; acquisition process for strategies, attitudes and styles; internalisation; transfer and sustainability. In the formative phase the purpose is revision and in the summative there is an objective assessment of whether the approach is meeting expectations. The next chapter discusses different kinds of testing instruments that may be used to construct items that assess critical thinking.

Watson, G. & Glaser, E. M. 1980, *Critical Thinking Appraisal Manual*, Harcourt Brace Jovanovich, New York.

An introduction to Glaser's test mentioned in other literature.

Wiggins, G. 1992, 'Creating tests worth taking', *Educational Leadership*, May 26–33.

This is a summary of principles for assessment task design. It reinforces the need to move beyond assessing recall of subject knowledge. Wiggins includes Resnick's criteria on which to judge tasks which aim to assess higher order thinking.

Wilen, W. W. & Phillips, J. A. 1995, 'Teaching critical thinking: A metacognitive approach', *Social Education* 593, 135–138.

This article proposes a metacognitive approach to the teaching of critical thinking skills in social studies. Links to active citizenship are established and the infusion method is preferred. The metacognitive aspect is added in terms of teacher and student modelling of critical thinking skills through reflection about the processes they employ to reach a conclusion or solve a problem.

Woods, D. R. 1983, 'Introducing explicit training in problem solving into our courses', *Higher Education Research and Development* 21.

Assuming that teachers want to introduce problem-solving strategies in their classroom, Woods sets up five stages through which the change can be implemented: evaluate the environment, define the problem, explore the options, develop a plan, implement and evaluate. There are connections to other theorists but Woods is concentrating on one aspect of the general understanding of critical thinking.

APPENDIX THREE:
FORM USED IN THE 2001 RANDOM SAMPLING EXERCISE

RANDOM SAMPLING OF YEAR 12 (2000) EXIT FOLIOS

PANEL'S CONSENSUS FORM

V H A	
H A	
S A	
L A	
V L A	

Please complete both sides of this form

CHECK that the details in the shaded box match with the school submission you have been given.

CONSIDER the materials in the school submission (work program, assessment instruments, student folios).

For each student, **DECIDE** a Level of Achievement and position in that achievement band.

RECORD, on the achievement band scale adjacent, the name of each student in the position corresponding to the decision you made for that student. The lower boundary of each achievement band represents the state-wide threshold for that band.

School name: «School_name»

Subject: «Subject_name»

Allocated district: «Area»

Selected students:

P : «Stud1»

Q : «Stud2»

R : «Stud3»

S : «Stud4»

T : «Stud5»

U : «Stud6»

V : «Stud7»

W : «Stud8»

X : «Stud9»

Signed _____ Panel Chair (or Nominee)

PLEASE ENSURE ONLY ONE CHOICE BUBBLE IS FILLED IN FOR EACH QUESTION.
FILL EACH BUBBLE DARKLY AND FULLY USING BLACK OR BLUE PEN.

Office Use

Please check whether the school submission is complete and it contains:

Yes	No	
○	○	a copy of the accredited work program
○	○	a set of all assessment instruments used in the school's determination of exit levels of achievement, with Forms R5 (or equivalent) and teacher's marking schemes/sample responses attached
○	○	all nine student folios
○	○	all responses to all assessment instruments for all students
○	○	completed student profiles.

Office Use ○ ○ ○ ○

Comment on any missing or additional material.

Reflect on the elements of the school submission and select the most appropriate response to the following statements.

	Strongly disagree	Disagree	Unsure	Agree	Strongly agree
All mandatory subject matter is covered and assessed by exit.	○	○	○	○	○
The assessment instruments reflect the intent of the syllabus.	○	○	○	○	○
The assessment instruments allow discrimination between students of differing abilities.	○	○	○	○	○
The grading/marking of student work is appropriate.	○	○	○	○	○
Global/overall standards for each assessment criterion (as shown on completed student profiles) reflect the relevant syllabus standards.	○	○	○	○	○
Sufficient information has been provided to enable decisions to be made about student achievement.	○	○	○	○	○

Comment on significant positive and/or negative aspects of the submission.

Office Use Only ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○

Reflect on how much the school's approach to assessment helps students develop thinking skills.

Does the school's approach to assessment:	Not at all	Only a little	To some extent	A lot
• encourage students to apply knowledge in demanding, unfamiliar situations (rather than rehearsed or routine situations)?	○	○	○	○
• give students sound opportunities to complete complex, open ended, multifaceted tasks (rather than overly structured tasks with too many cues)?	○	○	○	○
• allow students to be rewarded for demonstrating higher order thinking skills (such as thinking creatively, formulating and testing an hypothesis, using research material to analyse and synthesise, etc.)?	○	○	○	○

Please give reasons for your responses.

Draft