Criteria, standards and intuitions in the imprecise work of assessing writing

Jill Freiberg | Griffith University

Introduction
Standards-based assessment using explicitly stated criteria for both school-based and centrally organised common assessment is prescribed in current state and national programs that emphasise the need for comparability, consistency and accountability in curriculum at all levels of schooling. The application of generally applicable, pre-specified criteria and performance standards has been recommended as a procedure whereby the subjectivity of assessors’ qualitative judgements can be minimised and comparability achieved. It has been characterised as a value-neutral and efficient assessment technology.

An important part of the rhetoric of school-based, criteria- and standards-referenced assessment – apart from its reputation for certainty, objectivity and fairness – is its student-centredness. For both formative and summative assessment purposes, teacher-designed tasks and the pre-specification of criteria and standards are believed to eliminate the ill effects of other ‘subjective’ methodologies, because assessors’ judgements are based solely on comparisons between individuals’ performances and the explicitly stated performance criteria and standards. The assessment technology and procedures are given as insurance that assessment is objective, reliable, and therefore equitable, and that it will produce reliable assessments of levels of achievement.

Although criteria- and standards-based assessment has been widely used to assess written texts in senior secondary schools for twenty years, it is now to be routinely used throughout the early and middle years as well. It is timely, therefore, to revisit a review of research that has examined the ways that criteria and standards are applied during the assessment of students’ writing.

Writing and assessment
The production of written texts – either on paper or electronic media – is one of the most common tasks required of students for the assessment of literacy skills and procedures, and knowledge and understanding of language and textual resources. Writing is also the mode whereby knowledge and understanding and other assessable elements are presented for display in most curriculum areas. In this case, written texts are the carriers of displays of required learning and achievement. Students are required to compose posters, advertisements, PowerPoint presentations, newspaper articles, menus, health plans, design briefs, and many types of reports, recounts and other types of texts to display knowledge and understanding of facts, concepts, procedures and skills in investigation, reflection, analysis and synthesis of information and ideas. For the teacher, therefore, the crucial skill for assessing authentic achievement is the ability to differentiate between the carrier and the content, the medium and the message (McLuhan, 1964).
This paper focuses on this problem and reviews research on the criteria-based assessment of students’ written texts that, in accord with McLuhan (1964), demonstrates that the separation of form and content is impossible, even for skilled and experienced teachers. It is argued that, even when criteria and standards are explicitly specified, content knowledge and even literacy skills are so deeply embedded in the medium that the qualities of the medium influence not only their visibility to the assessor but also how they are perceived. The stated criteria for assessment become too difficult to apply directly; they must be re-interpreted and re-inflected by the assessor. As a result, it is the medium that is intuitively assessed, rather than the quality of its discrete features or the information it carries. Even the use of a supposedly objective assessment technology, such as standards-based assessment, simply disguises rather than eliminates this problem (Broadfoot, 1981).

The work of assessing

The role of the assessor of writing is much more than that of a ‘neutral technician’ making judgements that require a simple comparison between culture-neutral, ‘fixed’ standards and the assessable knowledge, skills and processes visible in the student text. In complex tasks, including tasks used to assess achievement in more than one curriculum area – where interpretation and inference is required of the assessor – the problem is intensified.

Assessors’ intuitions about the writer have been shown to play a large part in the valuing of and judgements about texts. These intuitions are gathered from many sources, including students’ classroom behaviour, the appearance of the text including surface features such as neatness and handwriting, and also the order and time of reading each text. A written text demonstrating above average knowledge and skills may be rated higher or lower, depending on the quality of the text read immediately before it. Based on such non-assessable characteristics, conditions and perceptions, the assessor positions the text as good or bad. For instance, in one early study, when the word ‘honours’ was randomly stamped at the top of students’ written texts, those texts received higher grades than others in the sample (Diedrich, 1974).

Many other factors have also been shown to influence the reliability of assessments of writing: assessor training procedures (Linn, Kiplinger, Chapman & LeMahieu, 1991; Herman, 1992), the level of monitoring during the assessment processes (Cooper & Odell, 1977), the rating techniques especially considering the speed of rating (Branthwaite, Trueman, Berrisford, 1981; McColly, 1979), whether rating was ‘holistic’ or ‘analytic’ (Freedman 1979), and assessor subjectivity (Freebody, 1990; Freiberg 2001; Gilbert, 1989; Hake & Williams, 1981; Ozolins, 1981). Although the effects of many of these can be reduced by changes in assessment procedures, assessor subjectivity has been found to be more intractable.

For example, Himley (1989) analysed six teachers’ (or assessors’) constructed readings of students’ writing in a group discussion. The participants’ talk showed that their judgements of the texts were influenced by, and constructed from, their intuitions about the attitudes, dispositions and character of the student writer. In fact, the primary task in the assessment process was the identification of the subjectivity of the student writer. The teachers noted that they ‘constructed a writer, a presence or “felt sense” in the text, and then rewarded “her”, albeit a bit ambiguously, with our communal stamp of approval’ “academic” she tried to be, the less we liked her’ (p.17). Evident in this is the seamless shift from judgements of the text to judgements of the person, and back again.

Such findings highlight the importance of ensuring that the pre-specified criteria are clearly stated and directly focused on the types of knowledges, skills and processes to be assessed and that they do not invite or require high degrees of interpretation and inference.
For instance, the application of criteria such as ‘evidence of a strong personal voice’, ‘originality’, ‘imagination and creativity’ require a great deal of inferential and interpretive work. The enactment of such criteria at the interface between the assessor and the text has been shown to be influenced both by subtle stylistic characteristics in texts and by the attitudes, dispositions and intuitions of the assessors.

For instance, in a study of the effect of ‘register’ in student essays, Freedman (1984) showed that, in looking for ‘personal voice’ in written texts, assessors made very fine distinctions of value based on their personal preferences and intuitions about the type of student whose writing was being assessed. Participant teachers were asked to assess student essays, some of which were written by students of various ‘abilities’ and some of which – although the assessors did not know it – had actually been written by adult, professional writers. The results of the study showed that there was a higher degree of assessor consensus about holistic ratings of real student essays than there was in the ratings of ‘fake’ student essays. In general, student essays were assessed more favourably. Analysis of the essays showed that differences between the professionals’ texts and students’ texts included length (the student essays were shorter) and degree of ‘formality’ (professionally written essays were more informal and more ‘familiar’ in tone, aiming to establish a close and yet authoritative relationship with readers). Strategies used by professional writers to achieve ‘informality’ included use of the first person pronoun, direct address to the reader and writing about ‘deeply held beliefs and novel ideas’. None of the strategies used by professional writers complied with the preferred relationship of the institutional assessor with the institutionally ‘appropriate’ subordinate and compliant student. Freedman noted that the student writing that was preferred used ‘linguistic forms that show respect, deference, and the proper degree of formality’ (p. 341).

Particular stylistic qualities in writing appear to endow texts with a special value that seems to transcend other textual variables. Criteria seem to be applied differentially, depending on perceptions of overall style. Appraisals of inappropriate ‘style’ seem to focus assessors on elements of ‘mechanics’ and ‘usage’, which – though they are not the major ‘problems’ in the texts – are more easily identified and explained and are, therefore, cited by assessors as justifications for the grade given (see Freiberg, 2001; Gilbert, 1987, 1989; Hake & Williams, 1981). Student writing that is not stylistically appealing focuses assessors on the ‘mechanics’ of the text.

**Criteria, standards or intuitions**

Assessor subjectivity has also been found to comprise situated interpretation of criteria, application of ‘gate-keeping’ criteria, ability to recognise particular displays of criteria, and individual stylistic preferences. Several significant studies of the assessment of writing in Australian schools have demonstrated the difficulty that even the most experienced teachers had in maintaining their focus on the application of discrete skills and knowledge during the assessment of students’ written texts.

In 1989, an evaluation of the Queensland Writing Task (Nuyen, 1990) was carried out after the initial administration of the test to all Queensland Year 12 students in that year. The aim was to evaluate the appropriateness of the Writing Task and the management of the test. The study also evaluated ‘the practicality of criteria-based holistic assessments’ and the appropriateness of the criteria and standards that were used to assess students’ written scripts (Nuyen 1990, pp. 1–2). The study included a large participant group (174 markers and 28 chief markers) of Queensland secondary English teachers from both State and non-State schools, who were recognised within their schools as expert teachers of English and as such had been nominated for selection as markers.
Nuyen (1990) found that, although markers and chief markers were consistent regarding their judgements about what was appropriate and effective in students’ written texts, these judgements were not limited to demonstrations of the application of the knowledge, understanding and skills listed in the assessment criteria. The criteria appeared to be interpreted subjectively by the markers so that their judgements could be based on choice of topic, type of narrative genre, relevance, length, register, creativity, originality, and overt expression of values or beliefs (e.g. religiosity, or persuasive exposition described as ‘sermonising’). Nuyen’s research showed, evident in the excerpt from the findings below, that many stylistic ‘indefinables’ were used as ‘gatekeeping’ criteria even when they were not explicitly pre-stated as assessment criteria.

Most [chief markers] were irritated by the same factors: factual or historical inaccuracy; religious sermons; offensive language; over-use of clichés. Again, as with the markers, all strove to be objective and cast aside their personal prejudices. However, there could be a danger of a ‘halo effect’ if too many of these responses were found in the same folders or in consecutive folders. (Nuyen, 1990, p.18)

Features of students’ written texts that appeared to have influenced markers’ overall impressions of ‘quality’ and their criteria-based holistic assessments, therefore, related not to writing skills but to the subject positions adopted by the student writers and the assessors’ judgements about the appropriateness of these.

Significant effects on teachers’ applications and inflections of specified criteria have been found when students’ written texts are valued for their levels of self-expression rather than displays of literacy skills and accuracy. Gilbert’s (1989) analyses of teacher and student talk about writing and the assessment of writing demonstrated that the teacher consistently judged the appropriateness of student writing based on evidence of ‘personal voice’, ‘creativity’, ‘originality’ and ‘spontaneity’, irrespective of the stated assessment criteria. She found that when the author was ‘visible’ to the assessor, texts could be read as holistic, quasi-literary utterances; when the author was not ‘visible’, texts were read as a set of component skills and mechanical devices (e.g. punctuation, spelling, grammar). Thus evidence of poor literacy ‘skills’ was sought to justify intuitive judgements of poor style.

Student writing – not only creative writing or story narrative productions, but also analytical exposition, factual reports and transactional writing – has been found to be valued for its degree of self-expression (Gilbert, 1987, 1989) and ‘flair’ and ‘style’ (Freebody, 1990; Freiberg, 2001; Ozolins, 1981), rather than for the displays of competence definable as literacy skills. Further, teachers or assessors have been found to attribute the writing skills necessary to portray ‘self’, ‘voice’, ‘flair’ and ‘style’ to the aptitude and dispositions of students rather than to the curriculum in which they engaged. Writing competencies were routinely attributed to innate or ‘inherited’ ability, and ‘performances’ of writing tasks that were read as being produced by diligently learned skills were not valued as highly.

Implicit in this is the assumption that there is a hierarchy of valued cultural attributes which students must display through their writing performances if they are to be assessed as writing quality texts. If these culturally defined attributes are not displayed, the students may be assessed as incompetent writers, irrespective of the accuracy of their writing. Despite the routine uses of criteria and standards for assessment of students’ writing, the features of written texts that counted as displays of these all important attributes appeared to be ‘transcendent devices’ (Freiberg, 2001) that resonated with the assessors.

One study (Freiberg, 2001) of 38 English teachers’ interpretations of criteria used to assess writing found that stated criteria were routinely re-cast so that ‘invisible’ qualities, transcendent and indefinable in nature, were what came to count as displays of ‘good writing’. Because these qualities are unexplained and not explicitly definable in textual
terms, it becomes impossible for students who do not possess the valued ‘style’, as a legacy from their life outside the classroom, to learn how to display it in written productions for assessment.

Because of their elusive nature, these indefinable, transcendent displays were found to be realised through the assessors’ recognition of particular qualities in students’ texts. The level of subjectivity of judgements about ‘confidence’, ‘thought’, ‘insight’, ‘maturity’ and ‘imagination’ as qualities within texts was evident, because they could only be produced by the participants in self-referencing terms: ‘I can understand it, they’ve actually given me something that I can read’; ‘I love reading this. It’s something that, I think, attracts you immediately. You, as a teacher just think, “I like the way that sounds”’; ‘It just comes down to like a gut feeling’ (Freiberg, 1993).

The analysis of the criteria used to assess written texts found that most required interpretation; they were not explicitly specified in ways that would allow them to be read as specific textual practices. The descriptions of different standards, therefore, were also unanchored: first, because the criteria to which they were linked were undefined, and second, because the differentiations between the particular unspecified attributes of texts were made in comparative, quantitative terms such as ‘great’ and ‘reasonable’, ‘broad’ and ‘fine’. The effect of this was to afford the assessor a warrant to ‘invent’ the articulation of the criteria and the standards.

Under these conditions, assessors are able to decide what will count as meeting the criteria and as meeting it ‘greatly’ or ‘reasonably’, ‘more’ or ‘less’. In the enactment of criteria-based assessment, therefore, the actual criteria and standards are not central to the process. Rather, it is teachers’ meanings that are central. The standard at which a student’s writing is assessed is based, therefore, on the resonance it produces within the assessor. If a text can be read as ‘literature’ or ‘good’ writing, particular assessment processes are called into play. These processes are different from the processes that are enacted when the student text is appraised as ‘bad’ writing.

Smith (1991) also addressed the issue of teachers as readers enacting criteria-based assessment in senior secondary English in Queensland schools. Smith’s (1991) analysis of the reading practices of five experienced English teachers was based on the premise that, as teachers enacted criteria-based assessment of student writing, they drew on various reading ‘repertoires’. These repertoires were found to shape the ways in which criteria were inflected in the ‘readings’ of student texts. Smith’s research challenges the assumption that the explicit statement of criteria and standards means that teachers’ reading practices also become explicit.

Smith identified:

1. that teachers’ ‘readings’ were influenced by non-specified criteria and a range of contextual factors which formed the interpretive resources of the teacher/reader;

2. that teachers/assessors habitually justify judgements by referring specifically to 
   *mechanics* such as spelling and punctuation, while it is clear that these textual errors cannot alone account for the grades awarded, and

3. that teachers/readers position texts in relation to other texts and in relation to other past and recent experiences of assessment, and that this positioning affects the ways in which the criteria and standards are inflected in each instance of assessment.

Smith concluded that each instance of criterion-based assessment in schools is ‘a complex interaction between the readers, the single text, and other like texts’ (Smith, 1991, p. 65).
and that the presupposition that criteria and standards are applied objectively and reliably is contradicted in the actual school-based practices.

Conclusion

Assessment of writing is very complex and imprecise. It becomes perhaps more difficult for assessors to ‘read’ student writing as text rather than as the display of the person-behind-the-text when summative assessment is school-based. The inclination of the reader to look for the ‘author behind the text’ is satisfied when the teacher assesses the writing of students s/he knows. This prior knowledge, or at least the assumption of prior knowledge about the person, helps to produce the ‘reading’ of the student text.

It is clear that if assessment of authentic achievement in writing is to be achieved, criteria that invoke creative and personal qualities such as ‘originality’ and ‘imagination’ must be explicated in terms of the knowledge and skills in writing technologies that are able to be taught and learned. Assumptions about assessor neutrality and the belief that assessment based on pre-specified criteria and standards is objective, reliable and equitable have allowed practices of assessment based on cultural displays – such as ‘style’, ‘sophistication’, ‘originality’, ‘imagination’, ‘maturity’ and development of ‘personal voice’ – to remain unexamined (Freebody, 1990).

In the assessment of writing, therefore, we must consider many factors: the types of criteria and standards specified; the level of interpretation and inference required; the visibility of the assessable elements and the complexity of the tasks and written productions, and the influence of the medium on the message to be assessed. In the end, it is the work of the assessor – at the interface between the assessor and the item being assessed – that is the point at which these essential elements will be either delivered or denied.

References


