

3 Writing assessment: Language, meaning and marking memoranda

Bonny Norton
University of British Columbia

Introduction

In this chapter, I critically examine the marking memoranda or scoring guides used in the assessment of writing skills in pre-university contexts in South Africa, the USA and Canada. The three marking memoranda (respectively) were or are currently being used in school-leaving Matriculation Examinations (South Africa), the Test of Written English (United States) and the Canadian Language Benchmarks Assessment (Canada). My familiarity with these marking memoranda is drawn not only from existing literature, but also from my own practical and research experience with these instruments (Norton Peirce 1990, 1991; Norton Peirce and Stewart 1997). It is important to note that the three memoranda are not necessarily representative of the multiple forms of writing assessment found in each of the three respective countries. Rather, these memoranda provide a window through which assumptions about language, meaning, writers and readers implicit in a variety of essay marking memoranda can be analyzed and critiqued.

What does a marking memorandum have to do with debates on fairness and ethics in language testing? In this article, I demonstrate that marking memoranda are not 'neutral', 'practical' artifacts; they are used to determine the life chances of individuals and assume theories of language and meaning that are rarely made explicit. If test developers and markers are to be ethically accountable (Hamp-Lyons 1997; Norton 1997; Shohamy 1993), we need to examine more closely the assumptions implicit in the marking memoranda we use, the decisions that we make on the basis of such memoranda, and the consequences these decisions have for the life chances of test takers.

It is not my intention to provide detailed analyses of the essay tests in which these memoranda are used. Readers are referred to other literature for this purpose (Educational Testing Service 1989; Greenberg 1986; Norton Peirce 1990, 1991; Norton Peirce and Stewart 1997; Raimes 1990; Stansfield 1986; Stansfield and Ross 1988). Furthermore, it is not my intention to assess the validity of the marking memoranda with respect to the purposes for which

the tests were constructed. My more modest intention is to determine what insights can be gained by a cross-comparison of marking memoranda used in the assessment of essay writing tasks – a very common task used in the assessment of writing skills. Comparative analysis is a useful methodological tool because the search for similarities and differences often uncovers hidden assumptions and implicit theories. In this spirit, the questions I will address in this chapter are as follows:

- 1 What are the similarities and differences among the three marking memoranda?
- 2 What assumptions about language and meaning are implicit in each of the memoranda?
- 3 How are writers and readers theorized – either implicitly or explicitly?

In the first section of the chapter, I compare the three essay-marking memoranda; in the second section, I provide a summary comment on the comparison and contrast, highlighting the relevance of the analysis for debates on fairness in language testing.

Comparison of three marking memoranda

In the interests of brevity, acronyms will be used for the three essay-marking memoranda. The one that was used in South Africa (at least during the *apartheid* years) will be referred to as the DEC – from the Department of Education and Culture; the one used in the United States will be referred to as the TWE (from the Test of Written English); and the one used in Canada will be referred to as the CLBA (from the Canadian Language Benchmarks Assessment). Copies of the three marking memoranda are available in Appendices A, B, and C in this chapter.

Central assumptions of the DEC marking memorandum

The DEC marking memorandum (Appendix A) was used in the marking of the essays written by black South African students in their national school-leaving examinations during the *apartheid* years. Essays were marked out of 70, on the basis of two ratings: a ‘language’ rating (see column 1) and a ‘content’ rating (see column 2). The highest mark, A, represented a mark of 56–70, and H, the lowest mark, represented a mark of 4–13. It has already been noted (Norton Peirce 1990) that a discrepancy existed between the language curriculum designed for black students, and the way student essays were actually marked in national examinations. The central assumptions about language and meaning – as reflected in the DEC marking memorandum – are as follows:

The theory of language implicit in the memorandum is that language represents – and is primarily limited to – knowledge of grammar, punctuation, vocabulary, and sentence structure, i.e. language is equivalent to linguistic competence and meaning is assumed to be contained within linguistic units. A clear distinction between ‘language’ (Column 1) and ‘content’ (Column 2) is drawn. As Norton Peirce (1990: 7) notes, markers are advised that: *‘The general principle applied is that the symbol to be awarded (A–H) is dependent on the use of language ... You must distinguish between major and minor errors.’* Norton Peirce notes that markers are given specific instructions to circle the major errors and underline the minor ones, and to take lexis and structure into account in determining what ‘language’ rating the writer should get. Once the writer’s ‘language’ rating has been determined, markers are then required to assess the ‘content’ of the essay. The quality of the content is determined with reference to the following questions (Norton Peirce 1990: 7):

- How well did the candidate relate to the topic?
- Is the topic introduced and concluded effectively?
- Does the essay hold the reader’s attention through interesting description, or imaginative writing, or perceptive ideas?
- Is it generally coherent?

Markers are warned not to become so distracted by the language that they forget the content. Rather, markers are advised to *‘read quickly over the essay again after marking the language’* (Norton Peirce 1990: 7).

It is important to notice that ‘language’, as depicted in the DEC memorandum, is significantly more important than ‘content’, when their respective weightings are considered. Thus once a writer’s ‘language’ has been pegged at a particular level, the marker has very little opportunity to exercise judgement about the writer’s ‘content’. For example, if a writer’s language has been deemed to be ‘patently below standard’ (14–22 out of 70), then regardless of how interesting the essay is, the marker may not give the writer a mark above 22, i.e. the marker has to remain within the 9-mark range from 14 to 22 (i.e. a 12% range).

A writer whose language is weak is deemed incapable of producing content that is both interesting and impressive. A ‘doubtful’ essay, for example, is assumed to have ‘dull’ content, and may, at best, arouse ‘some interest’. The corollary of this assumption, as demonstrated in the marking memorandum, is that a writer whose language is deemed ‘competent’ is incapable of producing ‘dull’ content. Thus a writer with a C symbol for language is assumed to produce either ‘ordinary’ or ‘interesting’ content.

In this memorandum, the weaknesses of writers are considered more important than their strengths. Although writers are ranked from A–H, the top of Columns 1 and 2 focuses attention on the H rating, after which the focus

shifts to G, F, E, D, C and (finally) B and A. Markers are not encouraged to 'bias for best' (Swain 1985).

Central assumptions of the TWE marking memorandum

The TWE is the essay component of the Test of English as a Foreign Language (TOEFL) used by over 30,000 institutions in the USA and Canada to assess the English proficiency of pre-university candidates whose native language is not English (Educational Testing Service 1989). The TWE marking memorandum (Appendix B) is a 6-point holistic scoring guide. The assumptions implicit in this marking memorandum are as follows.

Content and language – framed broadly as 'rhetoric' and 'syntax' respectively – are understood to be distinct, though related categories. A score of '4' for example, 'demonstrates minimal competence on both the rhetorical and syntactic levels.' A paper with a score of 3 'demonstrates minimal competence in writing, but remains flawed on either the rhetorical or syntactic level, or both.' Thus, in contrast to the DEC memorandum, the control of grammatical structures, syntactic variety, and word choice are assumed to be no more important than the organization, development, and effectiveness of the writing. A paper in the '6' category, for example:

- effectively addresses the writing task
- is well organized and well developed
- uses clearly appropriate details to support a thesis or illustrate ideas
- displays consistent facility in the use of language
- demonstrates syntactic variety and appropriate word choice

Meaning is assumed to be contained within the text, but is not confined to linguistic units – as is the case with the DEC marking memorandum. As in the DEC marking memorandum, however, readers are assumed to be outside the process of meaning construction. They are rarely invited to exercise judgement, but instead are expected to 'assign' scores to papers that 'demonstrate' competence, 'display' consistent facility, and 'address' the topic. In this view, a paper assumes a life of its own, independent of both writer and reader.

In contrast to the DEC memorandum, the strengths of writers are assumed to be more important than their weaknesses. Thus the memorandum states explicitly: *'Readers should focus on what the examinee does well.'* It states, furthermore, that a candidate with the highest score (6) is not expected to produce an error-free paper – the paper may have 'occasional' errors. In addition, the highest score is given prominence at the top of the guide, with the lowest score relegated to the bottom of the guide.

Assumptions of the CLBA Scoring Guide

The CLBA is a task-based assessment instrument used to place new adult immigrants in Canada in language programs appropriate for their level of competence in English (Norton Peirce and Stewart 1997). The test is divided into a Stage I (beginner) and a Stage II (intermediate). There is a writing test, a reading test, and a listening/speaking test in each stage. There are four writing tasks in each of the two stages, and the most challenging writing task is one in which writers are expected to express 'complex' ideas in English (Norton Peirce and Stewart 1997: 24). Each task is marked on a 4-point scale and is assessed with the help of a decision tree that distinguishes between 'primary objectives' and 'secondary objectives' of the task. The assumptions implicit in the marking memorandum (Appendix C) are described below.

In contrast to the DEC and TWE memoranda, where 'content' and 'language' are either divorced (the DEC) or integrated (the TWE), there is an assumption in the CLBA that the 'content' and 'expression of ideas' are of primary importance in written communication, while the syntax, spelling, mechanics, and lexis of the text are deemed to be of secondary importance. This is evident in the distinction between 'primary' and 'secondary' objectives.

In this memorandum, the relationship between the reader and the writer is of central importance in the construction of meaning. Thus meaning is not assumed to be contained within linguistic units (the DEC), nor is it assumed to be 'buried' in the text, to be ferreted out by the reader (the TWE). Unlike both the DEC and TWE marking memoranda, meaning is assumed to be co-constructed by both the writer and the reader. This relationship is captured in the 'primary objectives':

- The writer addresses the purpose of the task in paragraph form
- The writer demonstrates an appropriate sense of audience
- The writer organizes and expresses ideas clearly
- The reader can follow and understand the writer's message
- The reader is appropriately informed by the content of the task

There is an assumption that the marking of essays requires judgement on the part of the reader. While the DEC memorandum reduces marking to an accounting process in which major and minor errors are added up and recorded, and the TWE assumes that the worth of an essay is largely self-evident and simply needs to be 'discovered' by the reader, who then assigns an appropriate score, the CLBA foregrounds the importance of the reader's judgement in the assessment process, and provides a decision tree to facilitate the decision-making process.

Discussion

Having examined the assumptions underlying the three different marking memoranda, I would like to provide a summary of the central issues, and then discuss their relevance for fairness in language testing.

The marking memoranda have different assumptions about the relationship between language and meaning. The DEC takes the position that meaning can be separated from language, and that ideas are less important than the words that convey them. In this view, meaning is conveyed through language, but separate from language. The TWE takes the position that language and meaning have a complex interrelationship, but meaning is stable in written form. In this view, meaning is 'contained' within texts, and recoverable by the reader. The CLBA assumes that meaning and language have a complex interrelationship, but that meaning is not stable in written form. It is co-constructed by both reader and writer, and subject to change.

As an extension of 1, the three memoranda have contrasting approaches to the relationship between 'content' and 'language'. Although all three memoranda (implicitly or explicitly) draw distinctions between language and content, they frame them differently, and establish different priorities in their assessment. The DEC distinguishes between 'language' (error-free words and sentences) and 'content' (organization and ideas). Content is considered of secondary importance to language. The TWE distinguishes between 'rhetoric' and 'syntax' and generally considers them equally important. In the CLBA, appropriate 'content' is understood with reference to a given task in the context of a relationship between the writer and the reader, and is given primary importance in the assessment process. Syntax, spelling, mechanics, and vocabulary are considered to be of secondary importance.

- 3 It follows from 1 and 2 that the three marking memoranda have different assumptions about competent writers and readers. In the DEC marking memorandum, a competent writer is deemed to be someone who has few errors in grammar, spelling, and vocabulary; competent readers are people who effectively decode text at the level of the sentence. In the TWE, competent writers can communicate ideas effectively in written form; competent readers can decode and comprehend both rhetoric and syntax. In the CLBA, competent writers can express ideas clearly with reference to a given audience; competent readers can make appropriate judgements about the quality of the writer's message.

- 4 Points 1, 2 and 3 have important implications for fairness in language testing. The same essay marked in accordance with each of the three marking memoranda could be assessed very differently and, potentially, have vastly different consequences for the life chances of test takers. The central point is that theories of language, meaning, writers and readers are not abstract and divorced from the practical decisions that language testers, teachers, and administrators have to make on a daily basis. Such theories need to be made explicit, carefully examined, and rigorously defended on ethical grounds. Such scrutiny might shed some light on an intriguing question: are different theories of language equally fair?

References

- Clapham, C. and D. Corson (Eds.) (1997) *Language Testing and Assessment: Volume 7, The Encyclopedia of Language and Education*. Dordrecht: Kluwer Academic Publishers.
- Educational Testing Service (1989) *Test of Written English Guide*. Princeton, NJ: ETS.
- Greenberg, K. L. (1986) The development and validation of the TOEFL writing test: A discussion of TOEFL Research Reports 15 and 19. *TESOL Quarterly* 23, 3: 531–544.
- Hamp-Lyons, L. (1997) Ethics in language testing. In Clapham and Corson.
- Norton, B. (1997) Accountability in language assessment. In Clapham and Corson.
- Norton Peirce, B. (1990) Student writers, the DET syllabus, and matric marking: A critical evaluation. *ELTIC Reporter* 15: 3–11.
- Norton Peirce, B. (1991) Review of the TOEFL Test of Written English Scoring Guide. *TESOL Quarterly* 25: 159–163.
- Norton Peirce, B. and G. Stewart, (1997) The development of the Canadian Language Benchmarks Assessment. *TESL Canada Journal* 14: 17–31.
- Raimes, A. (1990) The TOEFL Test of Written English: Causes for concern. *TESOL Quarterly* 24: 427–442.
- Shohamy, E. (1993) The exercise of power and control in the rhetorics of testing. In Huhta, A., Sajavaara, K. and S. Takala, (Eds.), *Language Testing: New Openings*. University of Jyväskylä, Finland.
- Stansfield, C. (1986) A history of the Test of Written English: The developmental year. *Language Testing* 3: 224–234.
- Stansfield, C. and J. Ross, (1988) A long-term research agenda for the Test of Written English. *Language Testing* 5: 160–186.
- Swain, M. (1985) Large-scale communicative language testing: A case study. In Savignon, S. and M. Berns, (Eds.) *Initiatives in Communicative Language Teaching*. Reading, MA: Addison-Wesley: 185–201.

Appendix A

DEC Essay Memorandum Total possible mark = 70

Column 1 Language Rating		Column 2 Modified by Content Rating	
H. 4-13 (6-19%)	Extremely weak almost unintelligible	1. Extremely poor 2. Entirely dull 3. Arouse a faint interest	4-6 7-10 11-13
G. 14-22 (20-32%)	Patently below standard	1. Very poor 2. Dull 3. In measure	14-15 16-18 19-22
F. 23-27 (33-39%)	Doubtful	1. Dull 2. Some interest	23-24 25-27
E. 28-34 (40-49%)	Passworthy	1. Ordinary 2. Interesting in parts	28-30 31-34
D. 35-41 (50-59%)	A comfortable pass	1. Ordinary 2. Interesting	35-38 39-41
C. 42-48 (60-69%)	Competent	1. Ordinary 2. Interesting	42-44 45-48
B. 49-55 (70-79%)	Very competent and pleasing	1. Ordinary 2. Interesting 3. Interesting with flashes of impressiveness	49 50-53 54-55
A. 56-70 (80-100%)	Outstandingly good use of vocabulary and structures	1. Interesting 2. Interesting with flashes of impressiveness 3. Superb	56-60 61-63 64-70

Appendix B

Test of Written English (TWE) – Scoring Guide

TEST OF WRITTEN ENGLISH (TWE) SCORING GUIDE

Readers will assign scores based on the following scoring guide. Though examinees are asked to write on a specific topic, parts of the topic may be treated by implication. Readers should focus on what the examinee does well.

Scores

- 6** Clearly demonstrates competence in writing on both the rhetorical and syntactic levels, though it may have occasional errors.
A paper in the category
- is well organized and well developed;
 - efficiently addresses the writing task;
 - uses appropriate details to support a thesis or illustrate ideas;
 - shows unity, coherence and progression;
 - displays consistent facility in the use of language;
 - demonstrates syntactic variety and appropriate word choice.
- 5** Demonstrates competence in writing on both the rhetorical and syntactic levels, though it will have occasional errors.
A paper in this category
- is generally well organized and well developed, though it may have fewer details than a 6 paper;
 - may address some parts of the task more effectively than others;
 - shows unity, coherence and progression;
 - demonstrates some syntactic variety and range of vocabulary;
 - displays facility in language, though it may have more errors than does a 6 paper.
- 4** Demonstrates minimal competence in writing on both the rhetorical and syntactic levels.
A paper in this category
- is adequately organized;
 - addresses the writing topic adequately but may slight parts of the task;
 - uses some details to support a thesis or illustrate ideas;
 - demonstrates adequate but undistinguished or inconsistent facility with syntax and usage;
 - may contain some serious errors that occasionally obscure meaning.
- 3** Demonstrates some developing competence in writing, but it remains flawed on either the rhetorical or syntactic level or both.
A paper in this category may reveal one or more of the following weaknesses:
- inadequate organization or development;
 - failure to support or illustrate generalizations with appropriate or sufficient detail;
 - an accumulation of errors in a sentence structure and/or usage;
 - a noticeably inappropriate choice of words or word forms.
- 2** Suggests incompetence in writing
A paper in this category is flawed by one or more of the following weaknesses:
- failure to organize or develop;
 - little or no detail, or irrelevant specifics;
 - serious and frequent errors in usage or sentence structure
 - serious problems with focus.
- 1** Demonstrates incompetence in writing
A paper in this category will contain serious and persistent writing errors, may be illogical or incoherent, or may reveal the writer's inability to comprehend the question. A paper that is severely underdeveloped also falls into this category.

Papers that reject the assignment or fail to address the question in any way must be given to the Table Leader.
Papers that exhibit absolutely no response at all must be given to the Table Leader.

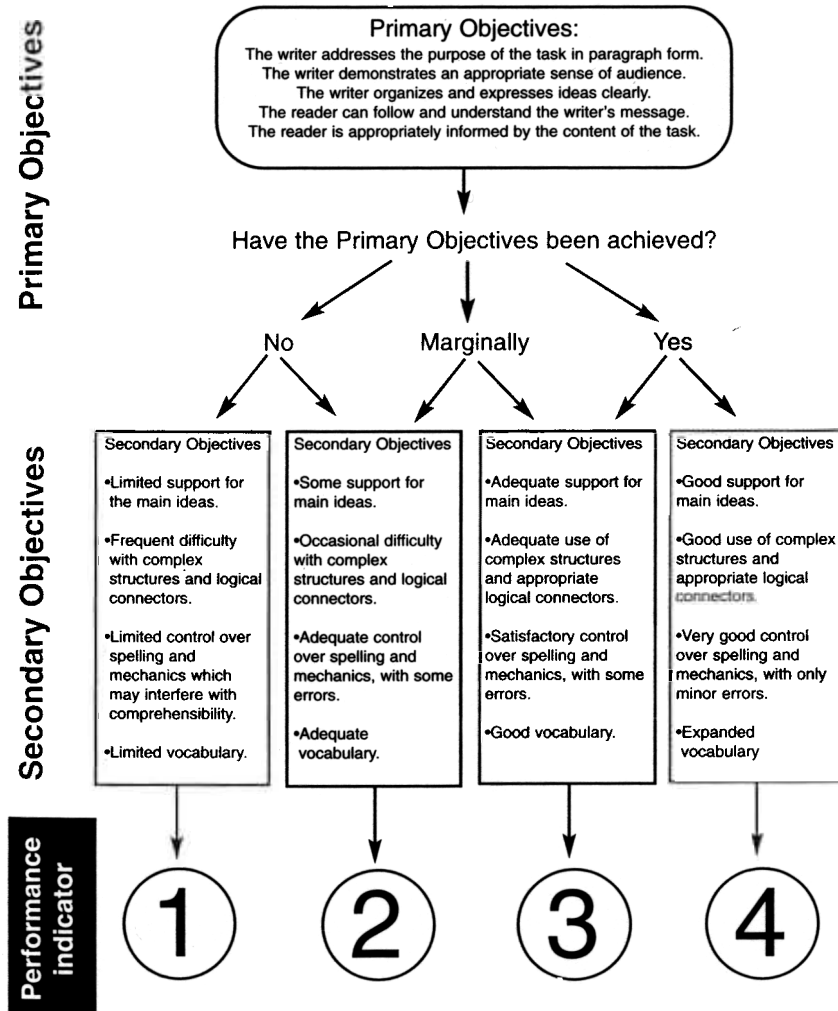
[Reprinted by permission of Education Testing Service, the copyright owner.]

Appendix C

CLBA Scoring Guide

CLBA Scoring Guide

Writing - Stage II - Task D:2: Express Complex Ideas



[Reprinted by permission of the Centre for Language Training and Assessment, Peel Board of Education, Canada.]