

ACER models of writing: Changing the assessment mindset

Juliette Mendelovits and Judy Nixon

Australian Council for Educational Research

<https://doi.org/10.37517/978-1-74286-685-7-4>

Juliette Mendelovits is the Research Director of the Assessment and Reporting research program at ACER. Since she joined ACER in 1991, Juliette has been involved in the conceptualisation, implementation and management of large-scale assessment projects, specialising in literacy, the humanities and social sciences. She played a leading role in the development of the OECD's Programme for International Student Assessment reading domain, from its first cycle in 2000 until 2012. Juliette has conceptualised and led the development of a number of reading and writing assessment projects at ACER, including the writing frameworks for ACER's International Schools' Assessment, Monitoring Trends in Educational Growth and the Southeast Asia Primary Learning Metrics. From 2015 to 2018, Juliette was based in London as Research Director and General Manager of ACER UK.

Judy Nixon is the Assistant Research Director of the Assessment and Reporting (Humanities and Social Sciences) research program at ACER. Since commencing at ACER in 2003, Judy has worked on the International Schools' Assessment, the Victorian General Achievement Test, the ACT Scaling Test, NAPLAN, NAP Civics and Citizenship, and a range of high-stakes selective tests and tertiary medical admissions tests. Judy has a special interest in writing assessment and research, and has managed the assessment of writing in a number of these tests including the creation of writing tasks, scoring rubrics, and marker training. She is currently part of a group working on a learning progression in writing.

Abstract

This presentation outlines different models of writing assessment that have a proven track record in various programs delivered by ACER. These models include holistic scoring, and partial analytical scoring on a number of criteria. The writing prompts for these models assume extended pieces of writing. The presentation will also introduce a new and alternative model of writing assessment that ACER has implemented in several large-scale assessments. This model is built on shorter pieces of writing from each student, designed to give more accurate and diverse insights into students' proficiency in writing in a range of text types, within the constraints of limited assessment time. Each short piece is coded using a different marking scheme with customised criteria. The three models – holistic scoring, partial analytical scoring and customised criterion scoring – are designed for different assessment contexts, from high-stakes selection to formative use by classroom teachers.

Introduction to ACER models of writing

ACER designs and delivers many writing assessments, from high-stakes assessments for selective schools and medical courses, to diagnostic assessments for classroom teachers working with their students. ACER currently uses three main forms of marking for this diverse range of tests: holistic marking, partial analytic scoring, and customised criterion scoring.

With holistic scoring, a student response is rated as a complete unit against a prepared scale. A global impression score is awarded denoting the level of achievement. Analytic scoring involves evaluating a student response by breaking it down into constituent elements or attributes, and assigning a mark for each criterion. Scores are aggregated to provide a single scale score. Two broad types of analytical scoring used by ACER are described here: partial analytical scoring, and customised criterion scoring.

Holistic marking

As noted, holistic marking at ACER involves assigning a global impression score to a piece of writing, usually on a scale from 1 to 10.

The ACER holistic model of writing from which a global impression is derived has three main assessment foci:

- Thought and Content (the quality of the thoughts, feelings and ideas presented)
- Structure and Organisation (the shape, form and cohesion of the piece)
- Expression and the technical aspects of language (the quality of the language used to present the ideas).

The three aspects in the model are treated somewhat hierarchically. Thought and content are prioritised over and above structure and the technical aspects of language in this model. This is because we posit that writing is about the process of generating ideas: creating ideas, not just recalling previously learned information. Writing can demonstrate original and divergent thinking, and the generation of new and varied responses to a set topic is valued. This process values writing as a type of *thinking*, through the written generation of ideas. This type of assessment of writing gives primacy to writing as a type of *thinking*, with the text's technical qualities of secondary importance. One advantage of this approach is that test takers with a language background other than English (LBOTE) are not penalised because of their lack of familiarity with the English language. Markers are always encouraged to 'look through' any obvious LBOTE errors in language, to the thought and ideas contained in the responses.

In the ACER tests of writing that use this scoring method, students may write in whatever genre or style they judge will best allow them to show their writing ability. Length is not a criterion. This form of scoring is used in the main for selective and admissions tests, for example the Cooperative Scholarship Testing Program (CSTP), the Graduate Medical Admissions Test (GAMSAT) and the Special Tertiary Admissions Test (STAT), where it is important to discriminate among applicants towards the top of the scale. Holistic scoring in these circumstances incorporates double-marking and discrepancy procedures, to ensure fair and accurate marking.

Partial analytic marking

ACER uses partial analytic marking schemas in some projects, both at school and tertiary level. It has been used in the International Schools' Assessment (ISA) test of writing since the test's inception. This is an assessment of writing, mathematical literacy and reading administered in over 335 international schools. It is undertaken by students from Grades 3 to 10 (pre-COVID, about 100,000 students in each annual cycle). It is underpinned by the principles of cultural eclecticism and administered by teachers in international schools in English, using detailed administration guides, but marked centrally by ACER markers. The participating schools receive reports at the school, class and student level and the students' results are used formatively to inform teaching and learning. Schools also use the assessment data to track progress over time and across year levels.

The writing component of the ISA comprises two extended writing tasks: one narrative/reflective task and one exposition/argument task. All students from Grades 3 to 10 write on the same two topics. The ISA Writing Task is based on historical development at ACER over several decades, of the concept of writing and the way students develop as writers. The marking and reporting scheme evolved for the ISA is very similar in nature to that used in the International Baccalaureate Middle Years Programme,¹ and the AERO (American Education Reaches Out) standards for writing.² The criteria for scoring also reference national examples of writing frameworks such as the McRel Standards,³ the Alberta writing program,⁴ the NWEA Six Traits,⁵ and the Australian Profiles for English and the New Zealand Curriculum.

For the ISA Narrative/Reflective task, students are asked to write a story or a reflective piece. The stimulus is usually a picture. The same prompt is used for all grades.

The ISA Exposition/Argument task requires a piece of writing that sets out ideas about a proposition. A few sentences or a short dialogue are provided as a prompt. Students may take an explanatory approach (exposition), a persuasive approach (argument), or they may combine the two approaches.

The ISA writing task uses a partial analytic scoring rubric. The criteria for assessing the Narrative/Reflective Task are:

- *Content* – broadly, the quality and range of ideas presented
- *Language* – sentence and paragraph structure, vocabulary and punctuation
- *Spelling* – phonetic and visual spelling patterns and the kind of words attempted, as well as correctness of spelling.

The criteria for assessing the Exposition/Argument Task are:

- *Content* – the depth and range of ideas presented, and the quality of reasoning
- *EAL Language* – the grammatical correctness and command of English syntax, sentence fluency and variation, and vocabulary
- *Structure and Organisation* – global and local organisation, and internal coherence.

Each student's writing thus receives six separate scores (that is, for three criteria on each of two tasks), which are then aggregated to give two scaled scores for writing. This approach to assessing writing is that it gives two kinds of reporting: a summative assessment of an individual's writing ability, which is mapped over time (from Years 3 to 10), and can be compared to others in the same year level both within schools and across schools; and important formative information for teachers to help them focus on their own teaching strategies.

Analytic scoring is also used in eWrite, an ACER online writing assessment which incorporates 10 different criteria across narrative, descriptive, informative and persuasive genres. In this assessment, because it is marked automatically using 'trained' algorithms, there is an emphasis on the technical aspects of student writing.

1 <https://www.ibo.org/programmes/middle-years-programme/curriculum/language-and-literature/>

2 http://projectaero.org/aero_standards/ELA/AERO-ELA-Framework.pdf

3 <https://www2.mcrel.org/compendium/topicsDetail.asp?topicsID=1013&subjectID=7>

4 <https://education.alberta.ca/media/160360/ela-pos-k-9.pdf>

5 <https://educationnorthwest.org/traits/traits-rubrics>

Customised criterion scoring

A third type of writing assessment design that ACER has pioneered in recent years is customised criterion scoring. The idea for developing this new form of writing assessment emerged from the recognised need to accurately assess what a student knows, understands and can do as a writer across a range of contexts. In more traditional forms of large-scale writing assessment, a very limited sample of a student's writing proficiency is collected: one extended piece of writing, or at most two, on which a judgement of proficiency is made. This means that the test taker has only one opportunity to demonstrate proficiency. Even when a partial analytical scoring model with several criteria is used – in some programs, such as NAPLAN writing, the single artefact is marked up to 10 times on specified criteria – the inescapable problem is that the same artefact is assessed repeatedly. The psychometric dependency among criteria is well-evidenced and probably unavoidable. We would not think of judging a student's maths or reading proficiency based on one item, would we? A further downstream problem of the single prompt instrument in a high-profile assessment is that it can have a stultifying effect on writing pedagogy, especially when a very limited and predictable range of text types is administered.

The key differentiating feature of the customised criterion scoring writing design is that each student is administered several writing tasks, within a manageable testing window. The assessment gives students the opportunity to demonstrate their proficiency across several different writing tasks, drawing on a range of writing skills both ideational and technical, covering different text types. Each task is marked on a different set of criteria drawn from an agreed framework but customised for the individual task. This is analogous to the way a reading or mathematics assessment is developed: each task reflects specific aspects of the framework.

The assessment framework for this kind of writing assessment comprises three dimensions or perspectives: content (text type – narrative, expository, instructional); context (the situation in which the writing is relevant, for example personal or 'wider world'); and process (e.g. generating ideas, structuring at micro and macro levels, syntax). Each writing task reflects one variable from each of the three dimensions. An important feature of the test design, or blueprint, is ensuring that, with regard to the process dimensions, there is a balance between the conceptual and the technical.

An individual student might be administered a set of three short tasks such as

- describing a picture in a couple of sentences (text type: descriptive)
- writing a quick note to a friend to arrange a play date, using details provided (text type: transactional)
- creating a half-page story based on a short textual prompt (text type: narrative).

Depending on the age and proficiency level of the students, at one end of the scale, a task might involve simply labelling a picture of an object (text type: label); at the other end, a task might ask the test taker to generate an argument of half to one page on a global issue (text type: persuasive).

Anything between two and five criteria are applied to a task, with the selection, number of scoring categories and specifics of the category descriptions customised to reflect the demands of the individual task. For example, a 'generate ideas' criterion with several scoring categories could be applied to each of the narrative and argument tasks sketched above, but would not be appropriate to the transactional or labelling tasks.

The multi-task writing assessment design, with customised criterion scoring, has been implemented by ACER in the New Zealand Literacy and Numeracy for Adults Assessment Tool (2010 to present), the Monitoring Trends in Educational Growth program (Afghanistan 2012) and in the regional assessment program, Southeast Asia Primary Learning Metrics (SEA-PLM 2019). This design is especially appropriate for sample assessment programs, in which each sampled student is administered a sample of the writing tasks, but it could also lend itself well to an adaptive writing instrument, using a combination of automated and human scoring.

Horses for courses

The three forms of writing assessment described all have their place. A single, holistic score captures an on-balance judgement about a piece of writing and has been shown, in analysis of programs including the writing assessments of the Western Australian Monitoring Standards in Education, and the Australian Cooperative Entry Program to be a reliable and efficient method of crystallising writing proficiency in a single score. It is appropriate where that is the main object, in selective entry programs such as those for scholarship selection or medical entry. Partial analytical scoring is suitable for combining the purposes of providing some formative analysis of students' proficiency in different aspects of the writer's craft, for teachers' use, with aggregated scaling of results for school or system consumption. Customised criterion scoring is particularly suitable for sample monitoring programs, and for reflecting a developed writing curriculum, where a wide range of text types and writing processes can be incorporated to give a comprehensive view of writing proficiency. Potentially, this kind of writing instrument is applicable to adaptive writing assessments, tailoring the type and difficulty of tasks to the proficiency of the individual test taker, and using a combination of automated scoring and human judgement.

In summary, each form of writing assessment and scoring has its virtues and deficiencies. A consideration of the purpose and context of the assessment will determine which type of writing instrument is most fit-for-purpose.