Construct and Content Validity

Test Validity
Test validity is an essential characteristic of a good test, but there are different kinds of validity, two of which are discussed below: construct validity and content validity.

Construct validity

do the definitions of the variable that is being tested and the way it is broken down into various categories and attributes match a common understanding of what the variable is?

The ISA assessment has three components: Mathematical Literacy, Reading and Writing.

The Mathematical Literacy and Reading components of the ISA are based on the constructs used for the Programme for International Student Assessment (PISA) which has been developed under the auspices of the Organisation for Economic Co-operation and Development (OECD). The definition and frameworks for the PISA domains of Mathematical Literacy and Reading were created by panels of international experts in the field, so there is strong international academic endorsement, by extension, for the concepts of Maths and Reading that are instantiated in the ISA.

The Writing component of the ISA is based on historical development at the Australian Council for Educational Research 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, and the AERO standards for writing – to name two international educational reference points – as well as with national examples of writing frameworks such as the McRel Standards, the Alberta writing program, the Six Traits (with the exception of Voice), the Australian Profiles for English and the New Zealand Curriculum.

Some more details about the constructs for each of the ISA domains follow.

Mathematical Literacy

The ISA adopts the PISA definition that ‘Mathematical literacy is an individual's capacity to formulate, employ and interpret mathematics in a variety of contexts. It includes reasoning mathematically and using mathematical concepts, procedures, facts and tools to describe, explain and predict phenomena. It assists individuals to recognise the role that mathematics plays in the world and to make the well-founded judgments and decisions needed by constructive, engaged and reflective citizens.’

In practice, Mathematical Literacy in the ISA is somewhat different from conventional mathematics, in that it is intended to encompass reasoning mathematically and using mathematical concepts, procedures, facts and tools in describing, predicting and explaining phenomena.

Each task in the Mathematical Literacy assessment is defined according to its content and the main type of process needed to complete it successfully.

Content variables

Uncertainty and Data This content area reflects how in real life data is commonly collected, organised, analysed and displayed with a view to making interpretations and forming conclusions. Many decisions are made based upon statistical analysis of data. Real life also contains elements of chance where outcomes are not certain but based upon probabilities. Increasingly decision-making is qualified with a statement of risk and society is presented with more and more information to make sense of, so competence in this area is of great significance in real life.

Quantity This overarching content area also features in the three other content areas to varying degrees. It focuses on the need for quantification in order to organise the world. It is not hard to find examples of quantification in our day-to-day living. We use money, make measurements, estimate and calculate. Increasingly we make use of technology to assist us but we also still perform many calculations mentally and approximately. Quantitative reasoning requires number sense: that is, having a feel for the magnitude of numbers, using strategies and tools appropriately, and being able to check solutions for reasonableness.
**Space and Shape**  Shapes and constructions are all around us physically as real objects but also as representations in the form of photographs, maps and diagrams. Constructing and interpreting such representations is an important skill. Using geometric shapes whose mathematical properties are known to model more complex shapes is an important problem-solving tool. Knowledge and appreciation of the beauty and function of geometric shapes and spaces has applications reaching from art to advertising.

**Change and Relationships**  Noticing and using patterns in number and shapes, and finding and describing relationships between variables, lies at the heart of mathematics. As organisms or populations grow and as stock markets ebb and flow, we describe the patterns in words, in tables and sometimes in algebraic notation. Commonly we chart the changes in graphical form. These patterns can be linear, non-linear, cyclic or exponential, to name but a few. Being able to link between these various representations and use the language, notation and algorithms of change and relationships is critical to making sense of the patterns in our world.

### Process variables

**Formulating** mathematics involves identifying opportunities to apply and use mathematics. It includes being able to take a situation as presented and transform it into a form amenable to mathematical treatment, providing mathematical structure and representations, identifying variables and making simplifying assumptions to help solve the problem or meet the challenge.

**Employing** mathematics involves applying mathematical reasoning and using mathematical concepts, procedures, facts and tools to derive a mathematical solution. It includes performing calculations, manipulating algebraic expressions and equations or other mathematical models, analysing information in a mathematical manner from mathematical diagrams and graphs, developing mathematical descriptions and explanations and using mathematical tools to solve problems.

**Interpreting** mathematics involves reflecting upon mathematical solutions or results and interpreting them in the context of a problem or challenge. It includes evaluating mathematical solutions or reasoning in relation to the context of the problem and determining whether the results are reasonable and make sense in the situation.

### Reading

Reading in the ISA is derived from the PISA concept of reading literacy, which was developed by an international panel of reading experts. Reading literacy in PISA is defined as ‘understanding, using and reflecting on written texts, in order to achieve one’s goals, to develop one’s knowledge and potential and to participate in society.’ While this definition and the construct of reading that grew out of it were developed with 15-year-olds in mind, the ISA construct of reading maintains the general thrust of a reading assessment that goes beyond the notion of decoding and literal comprehension (though at the lowest levels these are included), and recognises the full scope of situations in which reading plays a role for students from Grade 3 to Grade 10.

Each Reading task in the ISA is defined in terms of the *aspect* or approach to reading that it requires, and according to the *text format* of the reading passage on which the task is based.

### Aspect variables

**Retrieving Information** is defined as locating one or more pieces of information in a text.

**Interpreting** texts is defined as constructing meaning and drawing inferences from one or more parts of a text.

**Reflecting** is defined as relating a text to one’s experience, knowledge and ideas.

### Text format variables

**Continuous texts** are typically composed of sentences that are, in turn, organised into paragraphs. These may fit into even larger structures such as sections, chapters and books. Continuous texts include narrative pieces, exposition, description, argument and instructional passages.

**Non-continuous texts**, or documents as they are known in some approaches, can be described in structural terms as texts composed of one or more lists. In less formal terms, they can be described by their everyday appearance in such formats as tables, graphs, maps and diagrams.

### Writing

The ISA includes two extended writing tasks: one narrative task and one expository/argumentative task.

For the *Narrative* task the students are asked to write a story, either imaginary or based on experience. The stimulus may be a phrase, a sentence or a picture. The same prompt is used for all grades.
The **Exposition/Argument** task requires a piece of writing setting 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 same prompt is used for all grades.

In an effort to simulate good writing pedagogy, time is allowed at the beginning of each writing session for a brief class discussion of the topic, and for individual planning. Time is also allowed at the end of each session for students to proofread their work.

Students’ responses are evaluated on the basis of three criteria for each task.

### Criteria for assessing the Narrative Task

The **Content** criterion is about the quality and range of ideas presented, the development of plot, characters and setting, and the writer’s sense of audience and purpose. It also encompasses the overall shaping of the piece.

The **Language** criterion deals with sentence and paragraph structure, vocabulary and punctuation, and the writer’s voice.

The **Spelling** criterion takes into account students’ knowledge of phonetic and visual spelling patterns and the kind of words attempted, as well as correctness of spelling.

### Criteria for assessing the Exposition/Argument Task

The **Content** criterion looks at the depth and range of ideas presented, and at the quality of reasoning demonstrated in the ability to provide evidence and logical argument in support of a position.

The **ESOL Language** criterion is applied to all students’ writing regardless of their language background, but focuses on the grammatical correctness and command of English syntax, as well as sentence fluency and variation, and vocabulary.

The **Structure and Organisation** criterion deals with both global and local organisation: the overall structure of the writing, for example the presence of a clear introduction, development and conclusion; and its internal coherence, such as linking between and within paragraphs.

### Content validity

are the tasks in the test, including what the test-takers have to do (the test subject matter) and how they do it (the test format) – likely to give good indications of test-takers’ proficiency in the area being tested?

There are three salient features of the ISA to be remarked upon under this heading.

### Cultural content

A students’ performance in an assessment can be affected strongly by the familiarity of the contexts in which the test questions are set, even though the contexts are extrinsic to the skills or knowledge that the questions are designed to assess. This goes to the issue of fairness: it would, in the case of a test for international school students, be unfair as well as invalid if it catered mainly for students from one particular cultural background.

The ISA is written with international school students in mind, and therefore there is a strong emphasis in the selection of test material on catering for students from a wide variety of cultural and language backgrounds. The aim is to achieve cultural eclecticism, rather than cultural neutrality, so that all test takers will, ideally, find some familiar contexts and some unfamiliar ones in any given test.

ACER has a number of mechanisms for ensuring cultural appropriateness, including in-house reviews of all material, a panel of ‘critical friends’ – international school teachers and administrators who review all test material before it is selected, and responses to formal requests for feedback from teachers who have administered trial material.

### Task format

The Reading and Mathematical Literacy components of the ISA are in a combination of multiple-choice and constructed-response (short answer) formats, allowing for a wide range of skills, understandings and types of knowledge in the relevant areas to be interrogated. While factual knowledge and applications can be readily tested using closed test formats, it is often difficult to assess such skills as reflection and problem solving unless students are given the opportunity to generate their own responses. This underlies the decision to balance the task formats: it is a recognition of the constructivist pedagogy that is prevalent in international schools.

Writing is assessed by means of two extended responses to provided prompts. While it is not possible to simulate the writing process completely in a standardised testing context, some elements of that methodology are incorporated, by way of a brief class discussion of the writing topic at the beginning of the test session, and the opportunity to proof and edit at the end.
Level of difficulty of the test material

A valid test must aim to give accurate measurement of the ability of all the test takers, from the least able to the most able. If the test is too hard for the test takers, then we will have no idea of what they can do, only of what they can’t. If the test is too easy, then we will not be able to estimate the limit of the test takers’ proficiency.

In the ISA, the level of difficulty of the material is initially based on the test developers’ notion of what is appropriate content (and language) for each grade level. Since all test developers are former teachers, this estimate is a solid starting point, but it is then tested empirically during trial testing, when the actual difficulty of the tasks for the target group is observed through data collection and analysis of student responses. Item Response Theory (IRT), the method of statistical analysis used by ACER, allows the proficiency of the students and the difficulty of the tasks to be calibrated on the same scale. This technique enables us to select tests that match the range of proficiency levels of the target group, allowing us to construct tests that will measure with reasonable accuracy the proficiency of students at every level: the least proficient students (with some very easy items) as well as the most able students (with some very challenging items).

Test Reliability

Reliability is about whether the test is measuring the variable of interest in a consistent way, such that one could generalise about the result. Things that can undermine reliability include tests composed of items that measure many different things, so that the result does not tell anything very meaningful about any particular thing.

A simple summary of this kind of reliability is provided by the internal consistency statistic.

This figure shows the extent to which all the items in the test are measuring something similar. A figure of 0.8 indicates that 80% of the variation in the measures is related to the construct (what we are trying to measure), and 20% of the variation in the measures is related to variation of things other than the construct: or “noise” (termed “measurement error”). A figure of 0.8 is considered a good statistic for internal consistency.

The following tables show the internal consistency figures for Mathematical Literacy and Reading for 2002 to 2013.

Further information

Email isa@acer.edu.au
Phone +613 9277 5555 or
Website www.acer.edu.au/isa
Table 1: ISA Test Reliability By Year By Grade

<table>
<thead>
<tr>
<th>Year</th>
<th>Grade 3</th>
<th>Grade 4</th>
<th>Grade 5</th>
<th>Grade 6</th>
<th>Grade 7</th>
<th>Grade 8</th>
<th>Grade 9</th>
<th>Grade 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>2002</td>
<td>0.80</td>
<td>0.81</td>
<td>0.82</td>
<td>0.81</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2003</td>
<td>0.80</td>
<td>0.84</td>
<td>0.89</td>
<td>0.86</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2004</td>
<td>0.85</td>
<td>0.85</td>
<td>0.88</td>
<td>0.90</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005 (Oct.)</td>
<td>0.83</td>
<td>0.87</td>
<td>0.87</td>
<td>0.88</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005 (Feb.)</td>
<td>0.84</td>
<td>0.88</td>
<td>0.83</td>
<td>0.88</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2006 (Oct.)</td>
<td>0.83</td>
<td>0.82</td>
<td>0.82</td>
<td>0.84</td>
<td>0.86</td>
<td></td>
<td></td>
<td>0.88</td>
</tr>
<tr>
<td>2006 (Feb.)</td>
<td>0.82</td>
<td>0.82</td>
<td>0.85</td>
<td>0.83</td>
<td>0.85</td>
<td>0.84</td>
<td>0.86</td>
<td>0.82</td>
</tr>
<tr>
<td>2007 (Oct.)</td>
<td>0.83</td>
<td>0.84</td>
<td>0.85</td>
<td>0.87</td>
<td>0.86</td>
<td>0.86</td>
<td>0.88</td>
<td>0.84</td>
</tr>
<tr>
<td>2008 (Feb.)</td>
<td>0.83</td>
<td>0.86</td>
<td>0.86</td>
<td>0.85</td>
<td>0.87</td>
<td>0.87</td>
<td>0.89</td>
<td></td>
</tr>
<tr>
<td>2008 (Oct.)</td>
<td>0.83</td>
<td>0.84</td>
<td>0.85</td>
<td>0.83</td>
<td>0.84</td>
<td>0.86</td>
<td>0.87</td>
<td>0.88</td>
</tr>
<tr>
<td>2009 (Feb.)</td>
<td>0.83</td>
<td>0.86</td>
<td>0.86</td>
<td>0.85</td>
<td>0.87</td>
<td>0.87</td>
<td>0.89</td>
<td></td>
</tr>
<tr>
<td>2009 (Oct.)</td>
<td>0.80</td>
<td>0.83</td>
<td>0.84</td>
<td>0.86</td>
<td>0.87</td>
<td>0.86</td>
<td>0.85</td>
<td>0.89</td>
</tr>
<tr>
<td>2010 (Feb.)</td>
<td>0.82</td>
<td>0.85</td>
<td>0.85</td>
<td>0.87</td>
<td>0.88</td>
<td>0.87</td>
<td>0.87</td>
<td>0.87</td>
</tr>
<tr>
<td>2010 (Oct.)</td>
<td>0.84</td>
<td>0.84</td>
<td>0.86</td>
<td>0.86</td>
<td>0.86</td>
<td>0.86</td>
<td>0.86</td>
<td>0.88</td>
</tr>
<tr>
<td>2011 (Feb.)</td>
<td>0.86</td>
<td>0.86</td>
<td>0.87</td>
<td>0.88</td>
<td>0.89</td>
<td>0.87</td>
<td>0.88</td>
<td>0.88</td>
</tr>
<tr>
<td>2011 (Oct.)</td>
<td>0.85</td>
<td>0.87</td>
<td>0.84</td>
<td>0.84</td>
<td>0.84</td>
<td>0.85</td>
<td>0.85</td>
<td>0.86</td>
</tr>
<tr>
<td>2012 (Feb.)</td>
<td>0.86</td>
<td>0.87</td>
<td>0.86</td>
<td>0.85</td>
<td>0.86</td>
<td>0.86</td>
<td>0.87</td>
<td>0.88</td>
</tr>
<tr>
<td>2012 (Oct.)</td>
<td>0.84</td>
<td>0.84</td>
<td>0.83</td>
<td>0.86</td>
<td>0.87</td>
<td>0.86</td>
<td>0.85</td>
<td>0.88</td>
</tr>
<tr>
<td>2013 (Feb.)</td>
<td>0.87</td>
<td>0.85</td>
<td>0.86</td>
<td>0.87</td>
<td>0.88</td>
<td>0.87</td>
<td>0.87</td>
<td>0.88</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>StDev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.83</td>
<td>0.85</td>
</tr>
<tr>
<td>StDev</td>
<td>0.02</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Note: The data above represents the mean and standard deviation (stdev) of the reliability coefficients (Cronbach’s alpha) across different grades and years.
Table 2: ISA Test Reliability By Year By Grade

<table>
<thead>
<tr>
<th>Grade 3</th>
<th>Grade 4</th>
<th>Grade 5</th>
<th>Grade 6</th>
<th>Grade 7</th>
<th>Grade 8</th>
<th>Grade 9</th>
<th>Grade 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>2002</td>
<td>0.83</td>
<td>0.85</td>
<td>0.84</td>
<td>0.83</td>
<td>0.87</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2003</td>
<td>0.79</td>
<td>0.83</td>
<td>0.85</td>
<td>0.83</td>
<td>0.87</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2004</td>
<td>0.82</td>
<td>0.83</td>
<td>0.85</td>
<td>0.83</td>
<td>0.87</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005 (Oct.)</td>
<td>0.83</td>
<td>0.81</td>
<td>0.84</td>
<td>0.83</td>
<td>0.87</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2006 (Feb.)</td>
<td>0.83</td>
<td>0.84</td>
<td>0.86</td>
<td>0.83</td>
<td>0.87</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2006 (Oct.)</td>
<td>0.82</td>
<td>0.78</td>
<td>0.79</td>
<td>0.83</td>
<td>0.79</td>
<td>0.83</td>
<td></td>
</tr>
<tr>
<td>2007 (Oct.)</td>
<td>0.79</td>
<td>0.78</td>
<td>0.82</td>
<td>0.83</td>
<td>0.80</td>
<td>0.80</td>
<td>0.82</td>
</tr>
<tr>
<td>2008 (Feb.)</td>
<td>0.80</td>
<td>0.79</td>
<td>0.84</td>
<td>0.83</td>
<td>0.82</td>
<td>0.83</td>
<td>0.86</td>
</tr>
<tr>
<td>2008 (Oct.)</td>
<td>0.81</td>
<td>0.79</td>
<td>0.77</td>
<td>0.81</td>
<td>0.83</td>
<td>0.78</td>
<td>0.79</td>
</tr>
<tr>
<td>2009 (Feb.)</td>
<td>0.82</td>
<td>0.79</td>
<td>0.79</td>
<td>0.84</td>
<td>0.83</td>
<td>0.83</td>
<td>0.82</td>
</tr>
<tr>
<td>2009 (Oct.)</td>
<td>0.78</td>
<td>0.82</td>
<td>0.79</td>
<td>0.83</td>
<td>0.81</td>
<td>0.78</td>
<td>0.82</td>
</tr>
<tr>
<td>2010 (Feb.)</td>
<td>0.80</td>
<td>0.83</td>
<td>0.82</td>
<td>0.84</td>
<td>0.83</td>
<td>0.82</td>
<td>0.84</td>
</tr>
<tr>
<td>2010 (Oct.)</td>
<td>0.79</td>
<td>0.80</td>
<td>0.81</td>
<td>0.84</td>
<td>0.83</td>
<td>0.83</td>
<td>0.82</td>
</tr>
<tr>
<td>2011 (Feb.)</td>
<td>0.81</td>
<td>0.83</td>
<td>0.83</td>
<td>0.88</td>
<td>0.86</td>
<td>0.83</td>
<td>0.84</td>
</tr>
<tr>
<td>2011 (Oct.)</td>
<td>0.79</td>
<td>0.77</td>
<td>0.78</td>
<td>0.84</td>
<td>0.83</td>
<td>0.79</td>
<td>0.79</td>
</tr>
<tr>
<td>2012 (Feb.)</td>
<td>0.82</td>
<td>0.80</td>
<td>0.83</td>
<td>0.86</td>
<td>0.85</td>
<td>0.85</td>
<td>0.82</td>
</tr>
<tr>
<td>2012 (Oct.)</td>
<td>0.79</td>
<td>0.77</td>
<td>0.78</td>
<td>0.81</td>
<td>0.79</td>
<td>0.76</td>
<td>0.78</td>
</tr>
<tr>
<td>2013 (Feb.)</td>
<td>0.82</td>
<td>0.80</td>
<td>0.81</td>
<td>0.83</td>
<td>0.81</td>
<td>0.79</td>
<td>0.80</td>
</tr>
</tbody>
</table>

| mean    | 0.81 | 0.80 | 0.81 | 0.83 | 0.83 | 0.82 | 0.81 | 0.83 |
| stdev   | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.03 | 0.03 |

It can be seen that of the test instruments represented in tables 1 and 2 nearly all have good internal consistency (at or above 0.8). The exceptions are very close to the 0.8 benchmark.

Overall, the test reliabilities have means in the range of 0.80 and 0.87 from grade 3 to grade 10, which indicates that ISA tests have very good reliability in general for ISA mathematical literacy and reading from 2002 to 2012Oct/2013Feb. The standard deviations are in the range of 0.01 to 0.03, which means that the reliability values are consistently good.