

## e-asTTle Glossary

|                            |  |
|----------------------------|--|
| Achievement Objective (AO) | A specified part of the unpacked subject curriculum described in the e-asTTle <i>curriculum map</i> .  |
| Advanced (A)               | Student is consistently meeting the criteria at this level. Little disconfirming evidence is found. This student is ready to move on to material at the next <i>curriculum</i> level.  |
| Analytic scoring           | A method of subjective scoring, often used in the assessment of writing and speaking skills, where a separate score is awarded for each of a number of features of the task, as opposed to one global score (see <i>holistic scoring</i> ).                      |
| asTTle                     | Assessment Tools for Teaching and Learning.  |
| Basic (B)                  | Showing signs of these elements. Elements are evident in embryonic form. This is the entry level behaviour described by the curriculum for this level.   |
| Calibration                | Determining the value of a test <i>item</i> against a particular measurement scale – reflects <i>item</i> difficulty. IRT methods are used for this. (See also <i>item calibration</i> ).  |
| Cognitive processing       | The level of complexity required to respond to an <i>item</i> . e-asTTle uses a <i>SOLO</i> taxonomic description of an <i>item</i> . Two categories are used – <i>surface</i> (uni-structural and multistructural) and deep (relational and extended abstract). |
| Component                  | A high level description that groups related functionality together.   |

|                           |   |
|---------------------------|---|
| Construct                 | The trait or traits that a test is intended to measure. Can also be defined as an ability or set of abilities that will be reflected in test performance, and about which inferences can be made on the basis of test scores.   |
| Constructed response      | An <i>item</i> that requires the test taker to provide their own response. Sometimes called supply <i>items</i> , as the test taker has to supply the answer. For example, short answer, essay, cloze procedure, and performance assessments. (See also <i>selected response</i> ). |
| Content                   | The big ideas of the <i>curriculum map</i> . These may not correspond exactly with the strands in the related New Zealand Curriculum document.  |
| CTT                       | Classical Test Theory   |
| Curriculum level          | The levels specified in the New Zealand Curriculum that students should progress through as they move through their schooling, from Level 1 (entry) to Level 8 (at the end of Year 13).   |
| Curriculum map (e-asTTle) | An unpacking of the New Zealand Curriculum statement for a given subject by expert advisers.  |
| Cut score                 | The <i>logit</i> value that marks the boundary between two different levels. Arrived at as a result of the standard setting exercise.   |
| Deep                      | Relational and extended abstract <i>items</i> , when assessed using the <i>SOLO</i> taxonomy.   |

|                                     |   |
|-------------------------------------|---|
| Dichotomous                         | An <i>item</i> that is marked (scored) for one of two possible outcomes – T/F, Y/N or Correct/Incorrect. Marks/scores are awarded as 0 (incorrect) or 1 (correct).  |
| Differential item functioning (DIF) | A feature of an <i>item</i> that shows up in analysis as a group difference in the probability of answering that <i>item</i> correctly. The presence of differentially functioning <i>items</i> in a test has the effect of boosting or diminishing the total test score of one or another of the groups concerned. |
| Distractors                         | The supplied responses that are incorrect in multiple choice <i>items</i> .   |
| Domain                              | That portion of the total universe of subject matter that is being tested, and for which inferences can therefore be made.  |
| Estimates                           | The term used for measures of person ability and <i>item</i> parameters produced in latent trait models ( <i>IRT</i> ). These estimates are given in <i>logits</i> , units of measurement on a logarithmic scale.   |
| Factor analysis                     | A method of reducing the number of variables accounting for performance by identifying the underlying factor(s) shared by a set of test <i>items</i> . For example, a language test may have many <i>items</i> that have <i>items</i> related to the factors of listening, speaking, reading and writing.           |

|                  |   |
|------------------|---|
| Holistic scoring | A marking procedure that judges a piece of work (writing, speech, etc.) impressionistically according to its overall properties, rather than for the sum of its parts (see <i>analytic scoring</i> ).   |
| IEA              | International Association for the Evaluation of Educational Achievement. Based in Holland, and conducts several international benchmarking studies (e.g., PIRLS, TIMSS).  |
| IRT              | See <i>Item Response Theory</i> .   |
| Item             | <p>Osterlind (1990) offers this definition:</p> <p><i>“a unit of measurement with a stimulus and prescriptive form for answering; and, it is intended to yield a response from an examinee from which performance in some psychological construct (such as knowledge, ability, predisposition, or trait) may be inferred”</i> (p3)</p> <p>Essentially, it is a question/stimulus/prompt to which a student provides some response that can be scored, and from which we can determine their ability in the content area being assessed.</p> |
| Item bank        | A relatively large and accessible collection of <i>items</i> with known properties.   |
| Item calibration | The process of estimating the position of an <i>item</i> along a continuum (or line of the variable) along which persons are being measured. <i>Items</i> at one end of the continuum will be more difficult than those at the other end.   |

|                            |  |
|----------------------------|--|
| Item format                | A description of the different ways in which <i>items</i> can be constructed. For example, multiple choice, true/false, cloze- procedure.  |
| Item Response Theory (IRT) | Modern test theory that enables test <i>items</i> and students to be placed on the same scale of proficiency. Used in e-asTTle as the basis for calculating the <i>logits</i> for each <i>item</i> , and hence student and group proficiency for the different reports.                                    |
| Key (Answer)               | The correct answer for an <i>item</i> that will be awarded a score if present. It should include all possible variations and forms of acceptable answer.   |
| Key (Answer) rules         | The rules to apply for scoring an <i>item</i> . For example, “both required for one mark” where two responses are required.  |
| Link items                 | <i>Items</i> that occur in more than one trial paper and are used as comparison points for analysis purposes.  |
| Logit                      | An <i>IRT</i> statistic for each <i>item</i> that indicates the difficulty of the <i>item</i> . Used by the test compiler to select <i>items</i> for a test. Also used to determine a student score on the relevant scale. The usual range for <i>logits</i> is –3 to 3, but can occur outside this range. |
| Module                     | A high level description that groups a number of <i>components</i> into logical groupings.   |

|                    |   |
|--------------------|---|
| Polytomous         | <i>Items</i> that cannot be scored <i>dichotomously</i> (i.e., as a simple T/F or Y/N). <i>Items</i> where more than two values can be assigned as a score. Common on attitude and personality scales with multiple-response categories. In asTTle, when an <i>item</i> has a score value of 2 or more, then it is scored <i>polytomously</i> , as it is possible to obtain a score of 0, 1 or 2 (up to the maximum score value). |
| Proficient (P)     | There is evidence that the student is controlling or mastering the criteria elements. They should correctly answer <i>items</i> at this level about two-thirds of the time.   |
| RUMM               | Rasch Unidimensional Measurement Model. An analytical software tool. Uses <i>Item Response Theory (IRT)</i> to provide the underlying performance figures that drive the <i>item</i> selection and reporting functions of e-asTTle.   |
| Scope and Sequence | In many curricula around the world there is what is called a "Scope and Sequence". This is where certain standards need to be taught in certain order. For example, there may be a group of standards that are taught "pre-March" and another set "post-March". If the timeframe is "pre-March" then students should only be assessed on those standards.   |
| Score type         | Indicates whether an <i>item</i> is scored <i>dichotomously</i> or <i>polytomously</i> .  |
| Scoring            | The process of assigning a numerical value (score) to an answered <i>item</i> .   |

|                   |   |
|-------------------|---|
| Selected response | An <i>item</i> where the test taker has to select or choose the correct answer from a set of answers that are provided. For example, multiple choice. (See also <i>constructed response</i> ).  |
| SOLO              | Structure of Observed Learning Outcomes (SOLO). A cognitive processing taxonomy (classification system) devised by Biggs and Collis, and used in e-asTTle to identify <i>items</i> that are <i>surface</i> and <i>deep</i> in the <i>cognitive processing</i> required. |
| Standard setting  | The process of arriving at the <i>cut scores</i> that distinguish different levels of achievement within the curriculum.  |
| Stem              | The part of the <i>item</i> that asks the question or sets up the situation for response.   |
| Stimulus          | The material (text, diagram, graph, photo, etc) provided for which an <i>item</i> is written that requires a response from the test taker.  |
| Surface           | Uni-structural and multi-structural <i>items</i> , when assessed using the SOLO taxonomy.   |
| Technical reports | Documents that describe the research base of e-asTTle, and the reasons for the processes/actions taken.   |

## Testlet

Probably the most misunderstood term we use.

The term “testlet” was first used by Wainer & Kiely (1987), who defined it as “*an aggregation of items that are based on a single stimulus*”.

For instance, a reading comprehension test, which has a passage and a set of (say) four to twelve *items* that accompany the passage.

These *items* are not independent of each other, and therefore issues of misinterpretation, subject expertise, fatigue and so on are reasons for a test takers response to the *items* being more highly related to each other, than would occur with a set of totally independent *items*.

For e-asTTle, we use the term testlet to refer to the stimulus material that has several *items* that can appear together or independently.

Testlet images are stored separately from the question/*item*, so that the *items* can access the testlet image independently.

---

## Types of items

A description of the different ways in which *items* can be constructed. For example, multiple choice, true/false, cloze- procedure, short answer.