Assessment, Toulmin Arguments and Learning Progressions

Dr Kelvin Gregory
Senior Manager (Psychometrics)
with
Ms Rosita Rastan
Dr Somayeh Mahintorabi

Overview
The best schools in Australia are not necessarily those with the best ATAR or NAPLAN scores. They are those that enable their students to make the greatest progress in learning. Wherever a student starts from on the first day of the year, he or she deserves to have made at least a year’s worth of progress by the end of it. Any less, and our students will fail to reach their full potential. Sadly, that is too often the case.


Habermas argues that human thought and activity are divided into three categories:

- Work knowledge: objective world of facts that enables strategic action [instrumental knowledge]
- Practical knowledge: intersubjective world of community of actors with shared norms, roles, community values
- Emancipatory knowledge: subjective world of self knowledge, individual desires and feelings ['
  ‘collective emancipation from a history of domination that heretofore has come into being
  and proceeded spontaneously not guided by human reflection’ ”, Honneth & Joas, 1998:155 in Welton, p. 27)]

This presentation is largely in the technical or instrumental realm.
Illusory superiority and inferiority

Welcome to Lake Wobegon, where all the women are strong, all the men are good-looking, and all the children are above average.

Garrison Keillor

Illusory superiority is a cognitive bias whereby individuals overestimate their own qualities and abilities, relative to others.

In sum, the present study showed that illusory superiority at high school is beneficial for academic outcomes, while illusory inferiority is detrimental.

Wehrens, M.J.P.W., Kuyper, H., Buunk, A.P., & Van der Werf, M.P.C., 2008, Illusory superiority and inferiority at high school

Campbell’s law

Campbell’s law stipulates that "the more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor."

While largely technical it is important think at all levels ...

The government prescription of legally enforceable boundaries to school life limits what forms of education are possible, and individuality is suppressed by the monopoly of the management who make everything time, space, texts and procedures as uniform as possible (Gatto 2002: 78-79). Tensions occur regularly; the instruction of the class must go on even if particular students fall behind. This further demonstrates the effects of the colonisation of the life-world by the system: cognition, the basis of liberating education, is subjugated to the transferral of information.


Some fundamental questions of assessment

How do you define educational assessment?

What is the primary purpose of assessment?

How does assessment relate to the curriculum and course design?

What are the characteristics of good assessment? of poor assessment?

And there are many other questions that could be raised.
Defining assessment

Educational assessment is reasoning from observations of what learners do or make in a handful of particular circumstances, to what they know and can do more broadly (Mislevy, 2003, p.1).

Assessment is the systematic collection, review, and use of information about educational programs undertaken for the purpose of improving learning and development (Palomba & Banta, 1999, p. 4).

Assessment definition and purpose

Assessment is the process of gathering and interpreting evidence to make judgements about student learning. It is the crucial link between learning outcomes, content and teaching and learning activities. Assessment is used by learners and their teachers to decide where the learners are at in their learning, where they need to go, and how best to get there. **The purpose of assessment is to improve learning, inform teaching, help students achieve the highest standards they can and provide meaningful reports on students’ achievement.**

Misunderstandings about assessment

Assessment is only the test at the end of a unit of work or series of lessons.
Assessment is an auditing exercise about what students do and don't know or can and can't do, by testing student memory, asking trick questions etc.
Assessment focuses on what is easiest to measure.
Assessment is an average of performances across a teaching period.
Assessment is the same as grading.
Assessment for an achievement grade includes student dispositions and behaviours.


Common features of an assessment

Every assessment, regardless of its purpose, rests on three pillars:

► A model of how learners represent knowledge and develop competence in the subject domain
► Tasks or situations that allow one to observe learners’ performance
► An interpretation method for drawing inferences from the performance

Assessments are not developed and used in a value-free psychometric test-tube; they are virtually always intended to serve the needs of an educational system or society at large (Bachman, 1990, p. 279)

Roles for TIMSS, PIRLS, PISA, NAPLAN ... and ICAS?
Intended uses of large-scale assessments

- Monitor the quality of the education system for policy purposes
- Provide information for an evidence-based improvement cycle (e.g., In Japan)
- Provide data used to evaluate the achievement of equity goals
- Accountability purposes by reporting assessment outcomes to internal and external stakeholders
- Assessment may serve as policy levers

Roles for TIMSS, PIRLS, PISA, NAPLAN ... and ICAS?

Diagnostic power of assessment tool

Higher diagnostic power

4. Instruments that provide highly detailed extensive information on a student’s knowledge and understanding of narrowly defined skills or skill clusters

3. Instruments that provide information about individual’s strengths and weaknesses on narrowly defined domains or sub-domains

2. Instruments that provide limited diagnostic information but can alert teachers to potential problem areas e.g., NAPLAN

1. Frameworks describing levels of achievement

Lower diagnostic power
### Potential effects of high-stakes testing on learners

<table>
<thead>
<tr>
<th>Positive effects</th>
<th>Negative effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provide students with better information about their own knowledge and skills</td>
<td>Frustrate students and discourage them from trying</td>
</tr>
<tr>
<td>Motivate students to work harder in school</td>
<td>Make students more competitive</td>
</tr>
<tr>
<td>Send clearer signals to students about what to study</td>
<td>Cause students to devalue grades and school assessments</td>
</tr>
<tr>
<td>Help students associate personal effort with rewards</td>
<td></td>
</tr>
</tbody>
</table>

Do NAPLAN and ICAS have high-stakes effects on learners?

### Potential effects of high-stakes testing on teachers

<table>
<thead>
<tr>
<th>Positive effects</th>
<th>Negative effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Support better diagnosis of individual student needs</td>
<td>Encourage teachers to focus more on specific test content than on curriculum standards</td>
</tr>
<tr>
<td>Help teachers identify areas of strength and weakness in their curriculum</td>
<td>Lead teachers to engage in inappropriate test preparation</td>
</tr>
<tr>
<td>Help teachers identify content not mastered by students and redirect instruction</td>
<td>Devalue teachers sense of professional worth</td>
</tr>
</tbody>
</table>

Continued onto the next slide ...
## Potential effects of high-stakes testing on teachers

<table>
<thead>
<tr>
<th>Positive effects</th>
<th>Negative effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motivate teachers to work harder and smarter</td>
<td>Entice teachers to cheat when preparing or administering tests</td>
</tr>
<tr>
<td>Lead teachers to align instruction with standards</td>
<td></td>
</tr>
<tr>
<td>Encourage teachers to participate in professional development to improve instruction</td>
<td></td>
</tr>
<tr>
<td><strong>Do NAPLAN and ICAS have high-stakes effects on teachers?</strong></td>
<td></td>
</tr>
</tbody>
</table>

## Potential effects of high-stakes testing on administrators

<table>
<thead>
<tr>
<th>Positive effects</th>
<th>Negative effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cause administrators to examine school policies related to curriculum and instruction</td>
<td>Lead administrators to enact policies to increase test scores but not necessarily increase learning</td>
</tr>
<tr>
<td>Help administrators judge the quality of their programs</td>
<td>Cause administrators to reallocate resources to tested subjects at the expense of other subjects</td>
</tr>
<tr>
<td><strong>Do NAPLAN and ICAS have high-stakes effects on administrators?</strong></td>
<td><strong>Lead administrators to waste resources on test preparation</strong></td>
</tr>
<tr>
<td><strong>Lead administrators to change school policies to improve curriculum or instruction</strong></td>
<td><strong>Distract administrators from other school needs and problems</strong></td>
</tr>
<tr>
<td>Help administrators make better resource allocation decisions, e.g., provide professional development</td>
<td></td>
</tr>
</tbody>
</table>
"The fundamental purpose of assessment is to establish where learners are in the learning at the time of assessment.”
Masters, 2013

The key components:
- The data is the evidence used to prove something.
- The claim is what you are proving with the data.
- The warrant is the assumption or principle that connects the data to the claim.
The warrant links data to the claim

The warrant explains why the data supports the claim. They are the general, hypothetical (and often implicit) logical statements that serve as bridges between the claim and the data.

Rosita’s mathematics test score → Rosita is brilliant in mathematics
Learners who score high on the test are highly proficient in mathematics

Qualifiers, rebuttals and backing

- A qualifier is a statement about how strong the claim is. Qualifiers limit the strength of the argument or propose conditions under which the argument is true.
- Rebuttals are counter arguments or claims indicating circumstances when the general argument does not hold true.
- Backing statements are statements that support the warrant.
The complete Toulmin model

An application from Robert Mislevy (2003)

Pet Shop Display

Arturo is planning the parakeet display for his pet shop. He has five parakeets, Alice, Bob, Carla, Diwakar, and Etria. Each is a different color; not necessarily in the same order, they are white, speckled, green, blue, and yellow. Arturo has two cages. The top cage holds three birds, and the bottom cage holds two. The display must meet the following additional conditions:

- Alice is in the bottom cage.
- Bob is in the top cage and is not speckled.
- Carla cannot be in the same cage as the blue parakeet.
- Etria is green.
- The green parakeet and the speckled parakeet are in the same cage.

1. If Carla is in the top cage, which of the following must be true?
   a) The green parakeet is in the bottom cage.
   b) The speckled parakeet is in the bottom cage.
   c) Diwakar is in the top cage.
   d) Diwakar is in the bottom cage.
   e) The blue parakeet is in the top cage.
Applying Toulmin to Sue's response

W: Students who are high on Analytical Reasoning tend to do well on logical puzzles that query relations that follow from explicit relations and constraints.

B: Empirical studies show high correlations between AR test scores and college grades, open-ended problem solving tasks, and ratings of employees reasoning skills on the job.

C: Sue has a high value of Analytical Reasoning.

D1: Sue answered the Pet Shop item correctly.

D2: Logical structure and contents of Pet Shop item.

A: Sue answered correctly as a result of a lucky guess.

R: Sue spent less than 10 seconds on this item.

Imagine a test designed to assess a subtraction skill

\[
\begin{array}{c}
821 \\
-285 \\
\hline 536 \\
885 \\
-221 \\
\hline 664 \\
664 \\
63 \\
-15 \\
\hline 48 \\
17 \\
-9 \\
\hline 12
\end{array}
\]
And the resulting Toulmin method argument

**Diagram:**

- **W0:** Theory about how persons with configurations \((K_1, ..., K_m)\) would be likely to respond to items with different salient features.
- **C:** Sue’s probability of answering a Class 1 subtraction problem with borrowing is \(p_1\).
- **C:** Sue’s probability of answering a Class \(n\) subtraction problem with borrowing is \(p_n\).
- **W:** Sampling theory for items with since feature set defining Class 1.
- **D11:** Sue’s answer to Item j, Class 1.
- **D1n:** Sue’s answer to Item j, Class \(n\).
- **D21:** Structure and contents of Item j, Class 1.
- **D2n:** Structure and contents of Item j, Class \(n\).

---

**Summary of Toulmin and assessment arguments**

**Diagram:**

- **Decision to be made**
- **Warrants** since \(so\)
- **Backings**
- **Assessment-based interpretation** unless
- **Warrant** since \(so\)
- **Backings**
- **Assessment performance**
- **Assessment utilization argument**
- **Assessment validity argument**

---

**UNSW Global**
Defining learning progressions

Learning progressions are descriptions of:

- how learning typically advances in a subject area.
- successively more sophisticated ways of thinking about key disciplinary concepts and practices across multiple grades (Pellegrino, 2011, p. 9)
- the intermediate steps toward expertise (Pellegrino, 2011, p. 9)
- what it is that gets better when someone gets better at something (Leahy and Wiliam, 2011, p.1)
- how learners’ learning of important concepts and skills in a domain develops from its most rudimentary state through increasingly sophisticated states (Heritage, 2011, p.3)

An example from the USA’s Common Core Project

Grade 3
- Multiplications and divisions within 100
- Solve problems with the 4 operations
- Multi-digit arithmetic
- Understand fractions as numbers
- Solve problems involving time, liquid volumes, masses of objects
- Area, perimeter and operations

Grade 4
- Factors and multiples
- Generate and analyse patterns
- Multidigit arithmetic with whole numbers
- Fraction equivalence and ordering
- Decimal relations for fractions
- Measurement unit conversion from large to small
- Angles and measure angles
- Draw and classify shapes by properties of their lines and angles

Grade 5
- Foundations for multiplications
- Use place value to add and subtract
- Measure length in standard units
- Relate addition and subtraction to length
- Work with time and money

Grade 6
- Use ratios to solve problems
- Divide fractions by fractions
- Find common factors in multiples
- Rational numbers
- One-variable equations and inequalities
- Analyze relationships between HIs and DIs
- Real-world problems with area, surface, volume
- Statistical variability
- Distributions

Grade 7
- Use proportions to solve real-world problems
- Operations with rational numbers
- Generate equivalent expressions with operations
- Solve real-world problems using numerical and algebraic expressions
- Describe geometrical figures and relationships between them
- Solve real-world problems with angle measure
- Use random sampling to draw inferences about a population
- Understand and use probability

Grade 8
- Use rational numbers to approximate irrational
- Work with radicals and integer exponents
- Correct proportional relationships, lines, linear equations
- Solve pairs of linear equations
- Define, evaluate, compare functions
- Use functions to model relationships between quantities
- Understand and apply the Pythagorean Theorem
- Solve problems with cylinders, cones, spheres
- Investigate patterns of association in bivariate data
Empirical basis of learning progressions

"Empirically based learning progressions can visually and verbally articulate a hypothesis, or an anticipated path, of how student learning will typically move toward increased understanding over time with good instruction" (Hess, Kurizaki, and Holt, 2009).

- Learning progressions assume a progression from simple to more complex cognitive states
- Progression is neither random or linear
- Progression captures expected tendencies or likely probabilities

Learning progressions, grain size and assessment types

Learning progressions may be developed for virtually any grain-size, from multi-year curriculum frameworks to guides for minute-by-minute formative education/assessment.

Ideally, learning progressions should be multi-levelled and it should be possible to seamlessly change focus from one level to another

- Formative (assessment for learning) zooms in
- Summative assessments, including NAPLAN, zoom out and are suitable for monitoring progress over longer periods
Learning progressions, grain size and the class

Learning Progressions
Link the Zones of ALL Students

- Advanced
- Proficient
- "On track" for proficient
- Need for additional scaffolding?
- Modify materials & response formats?
- 2% Consistently Low Performing
- 1% Alt Assessment
- Many (but not all) students are here.
- What/how can they extend?

Learning progressions and the national curriculum - English

Large-grain learning progressions are built into the national curriculum

- Early stage 1: recognises some different purposes for writing and that own texts differ in various ways
- Stage 1: identifies how language use in their own writing differs according to their purpose, audience and subject matter
- Stage 2: identifies and uses language forms and features in their own writing appropriate to a range of purposes, audiences and contexts
Learning progressions and the national curriculum - mathematics

Learning progressions are built into the national curriculum

- Early stage 1: counts to 30, and orders, reads and represents numbers in the range 0 to 20
- Stage 1: applies place value, informally, to count, order, read and represent two- and three-digit numbers
- Stage 2: applies place value to order, read and represent numbers of up to five digit
- Stage 3: orders, reads and represents integers of any size and describes properties of whole numbers

Learning progression of fractions concepts and operations

<table>
<thead>
<tr>
<th>Year 7</th>
<th>Learners are solving multistep problems involving fractions or mixed numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year 6</td>
<td>Learners progress incrementally through multiplication and division of fractions</td>
</tr>
<tr>
<td>Year 5</td>
<td>Learners are able to add and subtract fractions and mixed numbers with unlike denominators</td>
</tr>
<tr>
<td>Year 4</td>
<td>Learners are able to add and subtract fractions with like denominators</td>
</tr>
<tr>
<td>Year 3</td>
<td>Learners develop an understanding of the meaning of a fraction</td>
</tr>
</tbody>
</table>
The ICAS suite of tests

ICAS is an independent skills-based assessment program. The program includes

- eight digital technology tests
- eleven English tests
- eleven mathematics tests
- eleven science tests
- six spelling tests
- ten writing tests
Building a numbers and arithmetic learning progression

If the same learner was given these ICAS mathematics assessments in successive years

▶ We could monitor their learning on an annual basis (large-grain size)
▶ and then make a claim about their progress

The grain size of an ICAS test is similar to that of NAPLAN but ICAS assessments are more frequent and do not have the same stakes associated with them.

Focusing on ICAS mathematics

ICAS mathematics tests are based upon an international framework as are other ICAS subjects (similar to TIMSS and PISA)

▶ Algebra and Patterns - Involves patterns of numbers, relationships between numbers and the use of symbols to stand for unknown or variable numbers.
▶ Chance and Data - Involves mathematical treatment of data and statistics.
▶ Measures and Units - Involves properties of the physical world that can be measured, the units used to measure them and the process of measurement.
▶ Number and Arithmetic - Involves types of numbers, their properties and number operations.
▶ Space and Geometry - Involves the properties of two-dimensional and three-dimensional space.
ICAS assessments have links to national curricula

<table>
<thead>
<tr>
<th>Year 3</th>
<th>Year 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count, order and compare whole numbers to 1000</td>
<td>Count, order and compare whole numbers to 10 000</td>
</tr>
<tr>
<td>Place value of whole numbers to 1000</td>
<td>Understand place value of whole numbers to 10 000</td>
</tr>
<tr>
<td>Skip by 2s, 3s, 5s and 10s; Use arrays to solve simple multiplication and division problems</td>
<td>Multiplication facts, of 2, 3, 5 and 10 and related division facts</td>
</tr>
<tr>
<td>Order and compare halves, quarters and eighths</td>
<td>Solve simple problems involving unit fractions with denominators 2, 3, 4 and 5</td>
</tr>
<tr>
<td>Solve simple addition and subtraction problems</td>
<td>Add and subtract to 100</td>
</tr>
</tbody>
</table>

Number and arithmetic items in past ICAS mathematics tests

These ICAS tests were focused on the competition component of ICAS. The past few years has seen ICAS move away from competition and awards to the assessment of learning.

<table>
<thead>
<tr>
<th>Year Level</th>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 in 2005</td>
<td>1 2 5 9 11 14 16 17 19 22 23 24 30 32 33 38 40</td>
</tr>
<tr>
<td>4 in 2006</td>
<td>2 3 6 9 12 14 16 18 19 21 22 25 28 30 31 36 38</td>
</tr>
<tr>
<td>5 in 2007</td>
<td>2 7 9 10 11 21 25 26 28 35</td>
</tr>
<tr>
<td>6 in 2008</td>
<td>3 4 7 11 16 19 23 30 31 35</td>
</tr>
</tbody>
</table>
Some technical considerations

It is challenging to compare learning from one term to another, and even more challenging to compare one year to another.

Rich electronic learning portfolios (which enable learner engagement) would be ideal but expensive to set up and maintain.

ICAS, NAPLAN and PISA all make use of item response theory, and specifically the Rasch model, to place test items onto scales that have invariant properties.

Growth can be measured providing certain "strong assumptions" are met.

A pilot project - the data

We extracted all the Australian ICAS mathematics data for

- 2005 for years 3 to 6
- 2006 for years 4 to 6
- 2007 for years 5 and 6
- 2008 for year 6

This covers core primary school years in all Australian states
A pilot project - merging

This data was then scored and merged to allow for
- vertical equating - placing all the test items with reference to year 3
- horizontal equating - placing all the test items onto the 2005 scale

Through this process all mathematics test items will be placed on the same number line

A pilot project - scaling

The Rasch model was used to place each item into the same scale. Specifically
- the Rasch model was used to model the probability of getting an item correct as a function of item difficulty and person ability
- a latent regression equation was used to take into account the different year levels

This is effectively using a model within the multidimensional random coefficients multinomial logit model used in TIMSS 1995 and in all PISA and NAP studies
Introducing logits

Item response theory models, including the Rasch model, work with logits.

Logit is short for "log-odds unit"

\[
\frac{\text{Probability of success}}{\text{Probability of failure}} = \text{ability} - \text{difficulty}
\]

Logits range from minus infinity to positive infinity

Logits obtained through the Rasch model have interval (measurement) properties (which the usual test scores lack)

Results of concurrent calibration
These ICAS tests were focused on the competition component of ICAS. The past few years has seen ICAS move away from competition and awards to the assessment of learning.

<table>
<thead>
<tr>
<th>Year Level</th>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 in 2005</td>
<td>1 2 3 4 5 29 30 31 32 33 34</td>
</tr>
<tr>
<td>4 in 2006</td>
<td>7 8 9 11 19 20 21 22 23 24 40 41 42 43 44 45</td>
</tr>
<tr>
<td>5 in 2007</td>
<td>15 16 17 18 21 23 25 26 36 39 40 41 43 45 46 48</td>
</tr>
<tr>
<td>6 in 2008</td>
<td>8 10 11 12 24 26 28 30 31 32 33 44 46 47 48</td>
</tr>
</tbody>
</table>
Some (free) alternatives

Some large scale assessment programs publish item response theory data and items for public usage. The following release a subset of items, all item parameters and describe learning progress benchmarks.

- Trends in International Mathematics and Science Studies (TIMSS) releases. Years 4, 8 and 12.
- Programme for International Student Assessment (PISA). Years 9-10.
- USA’s National Assessment of Educational Progress (NAEP). Years 4, 8 and 12.

You can:

- construct a benchmarking test using these released items.
- use the item parameters and learner’s response vectors to generate ability estimates
- transform the ability estimates onto the study’s reporting scale

And this applies to NAPLAN which has far more released items.
Some (free) alternatives

This can be very powerful, giving you an international perspective.

But you will still need to fill in the gap years.

Skill Maps

A skill map is the universe of skills available for learners to learn and the relationship among those skills to one another.

A skill map:

- is not learning progressions; learning progressions are specific components
- does not indicate whether or not learners have the choice of navigating from one skill to another
- reflect the skills that the learners is exposed to; it is the universe of skills
Elements of a skill map

The key considerations when developing a skill map:

- **Grain size**: the grain size of a skill to be learned (typically in a single lesson)
- **Relationships**: identify the prerequisite skills and create pathways

Core progress for mathematics (Renaissance Learning, 2013, p. 7)
Introducing diagnostic cognitive models

Item response theory models tend to be focused on summative assessments.
Cognitive diagnostic models (CDM) can be focused on summative and formative assessments.
CDM divide the latent space into fine-grained, often discrete or dichotomous cognitive skills or latent attributes, and evaluate the learner with respect to his/her level of competence for each attribute.

Types of skills

- Compensatory models: certain skills can compensate for the lack of others
- Non-compensatory models: the lack of a skill cannot be compensated for and the learner is likely to be unable to master the skill. Non-compensatory models are sometimes referred to conjunctive, meaning that all skills need to be mastered for a high probability of success on an item.
  - Some ICAS science items have a high reading load. Having a high reading ability will not compensate for a lack of science ability, nor will having a high science ability compensate for a lack of reading ability.
Cognitive diagnostic models features

Some defining characteristics of CDM

- multidimensional nature
- confirmatory nature
- complex loading structure
- criterion-referenced interpretations
- diagnostic nature of interpretations

Skill map resources

- https://www.turnonccmath.net/?p=map
Psychometrics

- A specialised field in education and psychology
- Focused on:
  - Constructing tools (tests, assessments, surveys, scales) to collect data
  - Developing, using and evaluating procedures to convert that data into measurements

Often what is being measured is latent or hidden and so much attention is paid to describing a construct.

Psychometrics in the Programme for International Student Assessment

You may be aware of the OECD’s Programme for International Student Assessment (PISA)
- Designed for 15 year olds
- Assesses literary, numeracy and other skills
- Sample based
  - Not designed for individual students
  - Designed for policy purposes
- Uses a specific modern measurement approach
  - The Rasch model

ICAS uses the same measurement approach and is designed to report at the student and other levels for a wider range of ages.
You may be aware of the Australia’s National Assessment Program (NAP, NAPLAN, NAPSL)

- Designed for years 3, 5, 7 and 9
- Assesses literary, numeracy and other skills
- Cohort based, stakes vary
  - Designed for individual students
- Uses a specific modern measurement approach
  - The Rasch model

ICAS uses the same measurement approach and is designed to report at the student levels for a wider range of ages.

The **construct** being measured must be carefully described.

- The measurement makes most sense when it is interpreted with reference to the construct description.
- The ICAS constructs are described in the Assessment Frameworks
  - And these descriptions are readily available from the web

The construct is conceptualized as a continuum from less of some quality to more of that quality.
The measurement continuum

High proficiency

Person ability

Low proficiency

Item difficulty

+2 Very hard item

+1

0

−1

−2 Very easy item

0.75 = \frac{e^{\theta-2}}{1+e^{\theta-2}}

\theta - 2 = 0.75
An ICAS mathematics item from 2008 (Year 3, Q19)

19. Mike uses this symbol ▲ for the number 23.

\[ 10 \times ▲ + 17 = \Box \]

What number is ▲ ?

(A) 27
(B) 170
(C) 247
(D) 400

The Rasch Model

The Rasch model, used in PISA, ICAS, Australia’s National Assessment Program, and many other assessment programs:

- Students answers are coded (or scored) as right or wrong
- These scores are then modelled as a function of student ability (\( \theta \)) and item difficulty (\( \beta \))

\[ P(X = 1) = \frac{e^{(\theta - \beta)}}{1 + e^{(\theta - \beta)}} \] (1)
Rasch model applied to an ICAS item

The Rasch model for Q19 in Year 3:

\[ p = 0.5 \]

\[ \beta = -0.71 \]

\[ \theta = -0.71 \]

We can say that a student with an ability of:

- -2.1 will have a 20% chance of getting the item correct
- -0.71 has a 50% chance of getting the item correct
- -0.01 will have a 67% chance of getting the item correct (PISA mastery level)
- 0.33 will have a 74% chance of getting the item correct (NAEP mastery level)

A general test design

High proficiency

Person ability

Low proficiency

\[ Q_1 \quad Q_2 \quad Q_3 \quad Q_4 \quad Q_5 \quad Q_6 \quad Q_7 \quad Q_8 \quad Q_9 \quad Q_{10} \quad Q_{11} \]
The person-item map

The person-item map from year 7 science
From logist to reporting scale

People tend to have problems interpreting logits.

TIMSS, PISA, NAP and ICAS all convert logits into a more 'friendly' metric. ICAS score = 1200 + 300*(1-logit)

Monitoring progress and growth

If you want to monitor learner progress and system change, then you need to:

- place assessments from the various years levels onto the same continuum
- and keep this continuum constant (stable) over assessment periods
  - Those periods could be weeks or months
  - and in ICAS they are years
Equating: the basic idea behind vertical equating using common items

Horizontal links: Keeping the scale constant
A snapshot: Preliminary vertical linking completed

Some key messages:

- The measurement model used in ICAS is the same as that used in PISA
- ICAS uses horizontal equating to enable comparisons across calendar years
  - Using common items, as does NAPLAN and PISA
- ICAS assesses 15 year old (year 9) students as does PISA
- ICAS assesses years 3, 5, 7 and 9 as does NAPLAN
  - ICAS assesses all years levels from 2 through to 12
  - And uses a rigorous equating system to do this

ICAS enables students and schools to benchmark growth and performance.
Our software

EAA/UNSW-Global psychometric team makes extensive use of in-house, commercial and open source software. Some of the programs used are:

- SAS
- SPSS
- Excel
- RUMM2030
- Conquest
- R
- \LaTeX