ACER General Ability Test New Edition

Assessment framework

Contents

March 2023

Figures and tables

Introduction

The ACER General Ability Test, commonly known as AGAT, is designed to measure the general reasoning ability of students through demonstrating an ability to identify relationships, process information and solve problems. The assessments have been developed especially, but not exclusively, for use in Australian schools and results can be compared to representative Australian norms at each year level. The AGAT construct is appropriate for broad international use and has been adapted to form the basis of a number of assessments developed by ACER for international contexts. AGAT has become a staple in schools' assessment programs, offering a test of general ability beyond what is covered in the curriculum.

AGAT 1st Edition has been available in paper format since 2008 and online since 2012.

AGAT 2nd Edition has been available since July 2022 and is only available online.

Note on terminology

Item definition: In this technical document, the word 'item' will be used throughout as this is more accurate than the colloquially used word 'question'. Test items are not always questions such as a statement, the instruction to 'complete this sentence' or the instruction to 'simplify' a mathematics equation.

AGAT 1st Edition

The AGAT 1st Edition assessments were the first assessments released in the AGAT series, but followed over 60 years of research and experience in the field of general ability testing. Initially released as paper-based assessments, AGAT was also made available as a suite of online assessments in 2012

AGAT 1st Edition comprises test forms ranging from Test 1 to Test 9 and can be administered according to student ability, based on previous scale scores and the educator's professional judgement. The test forms are targeted for students aged from seven to sixteen years of age.

From 2023, AGAT 1st Edition will only remain available for purchase as a paper-based assessment.

AGAT 2nd Edition

The AGAT 2nd Edition (2022) assessments are the most recently developed assessments to use the AGAT construct and they are only available via ACER's Online Assessment and Reporting System (OARS). Updated in line with the contemporary needs of educators and students, AGAT 2nd Edition continues to offer robust measures of general ability with an updated, colourful and engaging format.

AGAT 2nd Edition comprises test forms ranging from Test 1 to Test 9 and can be administered according to student ability, based on previous scale scores and the educator's professional judgement. The test forms are targeted for students aged from seven to sixteen years of age.

There are four types of standard reports issued in your online school account that provide details of your students' performance in AGAT. You can find guidance on how to generate these reports in the Help section of your account.

Rationale for AGAT

The ability to understand and analyse new information and apply prior knowledge to solve new problems is essential for everyday life. There are many students in schools that possess a level of general ability that may not be identified in achievement assessments or traditional learning areas. In cases where a student's reasoning abilities are not identified the student may miss out on opportunities for targeted support that would assist in raising achievement levels. Conversely, other students may need support in learning how to analyse and understand concepts and apply prior knowledge to solve problems. Using data from the AGAT assessments can support and inform teaching practices and be a part of a program of action.

AGAT 2nd Edition is a series of assessments designed to assist teachers in their assessment of students' general reasoning ability and support students' learning by identifying strengths and weaknesses and monitor growth in general reasoning ability over time.

AGAT reporting enables this by providing:

- scale scores, representing general ability points on the AGAT scale that allow results to be compared with past or future AGAT test results to monitor student progress
- qualitative descriptions of general ability levels based on the items in the tests
- Year (Grade) level percentile ranks from a comparison with the AGAT 2019 reference sample
- an overview of the student's strengths and weaknesses in each cognitive strand covered.

ACER's expertise in assessing general ability

ACER is one of the world's leading organisations in educational research, measurement and test development. For over 70 years ACER has been developing assessment frameworks and instruments for reliably measuring general ability and reasoning skills. In the 1940s ACER commenced developing the A.C.E.R. Adult, Advanced, Higher, Intermediate, Lower and Junior Tests to support vocation guidance in secondary schools and to provide robust measures of general ability for learners from the age of 8 years to adulthood. These tests were trialled with large cohorts of students and were published with standardised scores and normed percentiles available based on the findings from the trials and norm setting studies. Items and test forms were analysed quantitatively for validity and reliability, which over time provided ACER with a deep understanding in the measurement of general ability.

At the commencement of this century, ACER published the Middle Years Ability Tests (MYAT) and the Aptitude Profile Test Series (APTS), which built on the previous fifty years of ACER's research and fine tuning of general ability assessments. This included providing new and more engaging material and introducing equated test forms and scales. As a part of their development, the APTS and MYAT were trialled and analysed by ACER researchers, and norm studies were also conducted. The first edition of AGAT was a fully revised and expanded version of MYAT with some content being drawn from the APTS.

In the last five years, ACER has delivered AGAT assessment content in international schools through ACER's International Benchmark Tests (IBT) program. This has demonstrated the application of ACER's general ability constructs outside of Australia and across different cultures. The international administration also led to the expansion of the AGAT construct to include two new strands. The AGAT 2nd Edition items have been selected as the most reliable and valid items from item banks that have been developed, trialled, reviewed and approved for use between 2007 and 2021. These items were then trialled in schools across Australia to determine the appropriateness of their inclusion in AGAT 2nd Edition.

AGAT and the curricula

AGAT items are not developed according to the Australian Curriculum or any other curricula. This means that students do not need to have achieved any particular level in their school education to sit any level of the AGAT forms. Students can sit any test level that the classroom teacher thinks will be appropriate for that student's reasoning and problem-solving abilities.

Construct

Definition

In AGAT, the skills assessed require students to reason, make logical deductions, identify connections and spot patterns. The items do not require any prior knowledge from any particular content area. Instead, students must demonstrate their ability to think critically and use their higher order thinking skills to solve simple, multi-step and non-routine problems. The AGAT scale also recognises general ability is a continuum, whereby students develop their information processing and problem-solving skills at different rates. The AGAT levels that are derived from the AGAT scale also describe general ability at each level and collectively describe a progression of general ability.

Structure

It is critical to the construct of the AGAT assessments that items cannot be answered by simple recall or rote application of practiced skills. Ability assessments may be highly specific where a well-defined set of reasoning skills and contexts are assessed. For example, a verbal reasoning assessment will measure a person's reasoning skills in the domain of using words. This may include discriminating between words, accurately using different words or drawing a logical conclusion from a set of statements. Further examples are described in the next section.

A 'general' ability assessment uses a variety of types of reasoning skills and contexts, that collectively provide a measure of 'general' reasoning ability. In AGAT these categorisations of reasoning skills and contexts are known as 'cognitive strands'. An advantage of a general ability assessment is that in addition to measuring a student's general ability, the assessment also provides valuable information about the student's ability in each of the cognitive strands.

In AGAT 1st Edition, three cognitive strands were identified and assessed: verbal, numerical and abstract. In AGAT 2nd Edition, we have expanded this to five strands to include kinetic reasoning and spatial reasoning.

The inclusion of the two new strands makes the tests more interesting and provides richer information about areas of ability without compromising the validity or reliability of the general ability measures.

Strands

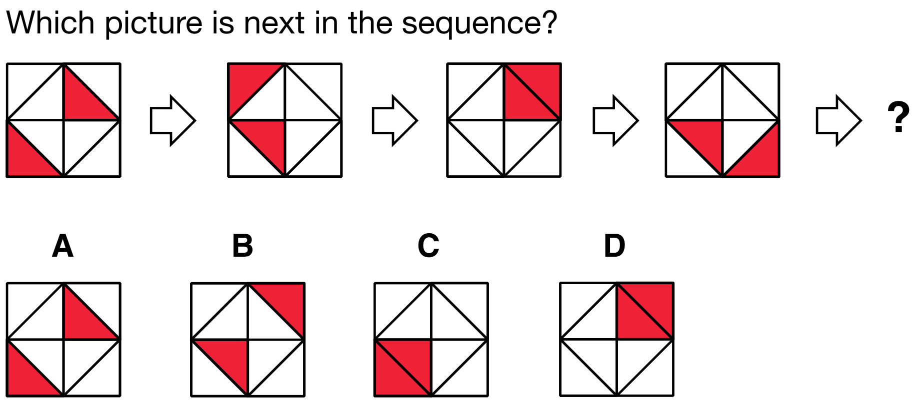
Abstract reasoning

Abstract reasoning is the ability to see patterns and logic in pictures and diagrams. Abstract reasoning items require students to complete visual patterns that follow simple rules, deduce which rules have been applied to change the states of images, and identify the next steps in 2-dimensional visual sequences. Abstract reasoning is decontextualised in the sense that the problems addressed cannot be applied to any real-world context – they deal with abstract concepts.

At lower levels, students may be required to spot simple rotational patterns in 2-D images, and identify the next step in that sequence for example. At higher levels, students may need to identify the rules that have been applied to transform multi-faceted shapes and apply those rules to new scenarios.

An example of an abstract reasoning item is shown. In this item, students have to follow the patterns created by the two rotating red triangles. The red triangle on the outer edge rotates one quarter turn clockwise while the inner red triangle rotates half a turn. Thus, the correct answer is A.

Figure 1 *Abstract reasoning example item*



Kinetic reasoning

Kinetic reasoning is the ability to anticipate the results from the movement of objects. Kinetic reasoning items require students to recognise the effects of turning gears, pulling levers and manipulating pulleys. They require students to understand the flow of water and rolling balls in simple networks, and the position of objects on a grid after a series of commands.

At lower levels, students may be required to identify what happens when a force is applied to a lever in a simple system. At higher levels, students may need to apply rules by tracking backwards to establish the starting point of a dynamic situation.

An example of a kinetic reasoning item is shown. In this item, students have to visualise the movement of the four drones on the grid to see which squares can be reached using the given rules. No drone can reach S4 in any one move, so Option D is the correct answer.

Figure 2 *Kinetic reasoning example item*

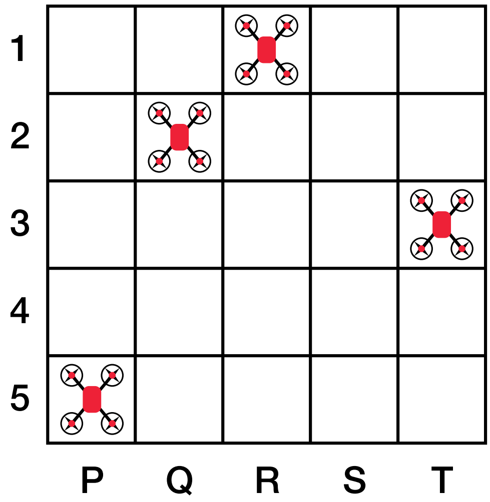
Drones on a grid can only travel left, right, forwards or backwards (not diagonally).

1 move = travel a maximum of 4 squares in one direction.

Four drones are shown in their starting positions on this 5 × 5 grid.

Which square can not be reached by any drone in 1 move?

A P1
B R3
C R4
D S4



Numerical reasoning

Numerical reasoning is the ability to use numbers to solve a variety of multi-step problems. Numerical reasoning items require students to recognise numerical patterns and sequences, categorise objects to match numerical quotas, link input and output from number machines, and apply rules to arithmetic puzzles.

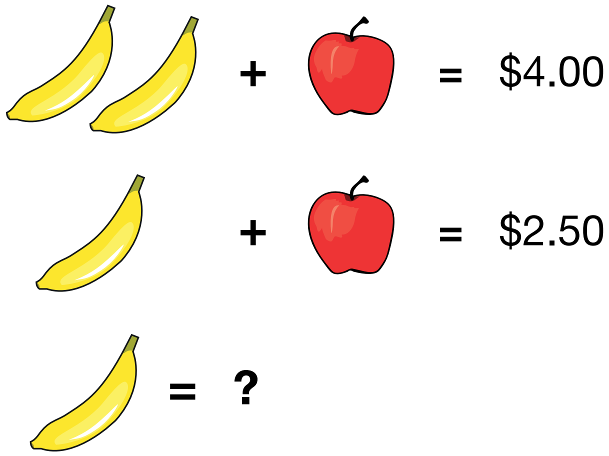
For example, at lower levels, students may be required to apply basic numerical deduction to calculate unknown values in simple word problems. At upper levels, students may need to consider multiple inter-related variables to find the outcomes of non-standard scenarios.

An example of a numerical reasoning item is shown below. In this item, students must find the value of a single banana. This problem can be solved using different methods (including using a simultaneous equation, noticing that the difference between the equations is the value of one banana or by plugging numbers in) to establish that a banana costs $1.50 (and an apple costs $1).

Figure 3 *Numerical reasoning example item*

How much does one banana cost?

A $0.50
B $1.00
C $1.50
D $2.00



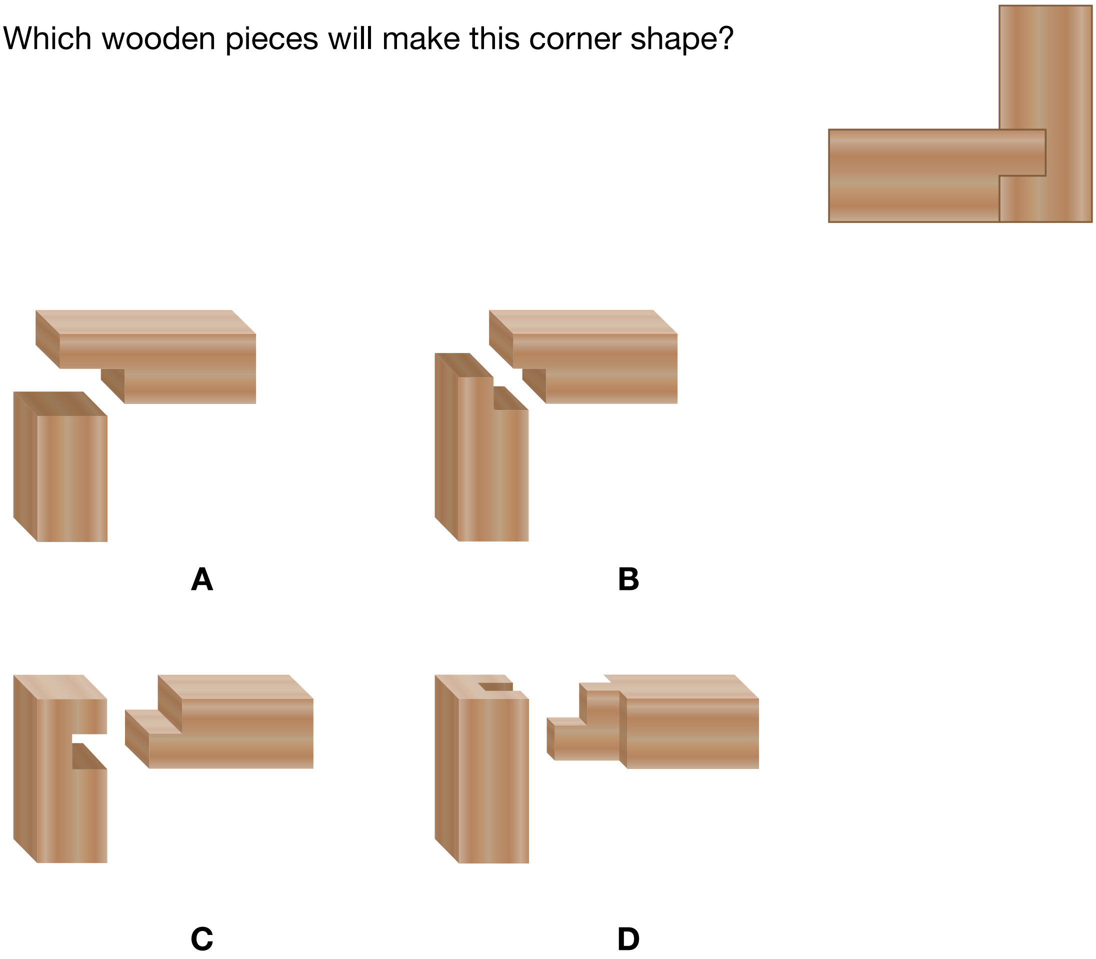
Spatial reasoning

Spatial reasoning is the ability to visualise the transformations of objects on a page. Spatial reasoning items require students to identify different viewpoints when looking at 3-D objects, recognise where shapes appear in complex images, identify how shapes have been manipulated through reflection and rotation, and rearrange pieces of an image to form a complete picture.

For example, at lower levels, students may be required to identify how to rotate two simple objects to make them fit together. At higher levels, students may need to recognise how a set of objects has been manipulated between photographs taken from different perspectives.

An example of a spatial reasoning item is shown. In this item, students need to recognise that the wooden blocks in Option C can slot together and be rotated to form the corner shape shown.

Figure 4 *Spatial reasoning example item*



Verbal reasoning

Verbal reasoning is the ability to understand how words connect to each other and how words within a sentence affect meaning. Verbal reasoning items require students to understand the hierarchy of words, identify relationships between words, rearrange words to form a sentence, and make logic of competing sentences.

For example, at lower levels, students may need to identify specific and general words from a group of similar words. At higher levels, students may be required to consider multiple sentences providing related information to specify the order in which things can be organised.

An example of a verbal reasoning item is shown below. In this item, students need to rearrange the nine given words into a sentence and identify which word will be at the end of that sentence. The rearranged sentence is: 'it is unusual to see so many happy faces', so Option A is the correct answer.

Figure 5 *Verbal reasoning example item*

Re-arrange the words to make a sentence.

many unusual faces see it is so happy to

Which word goes at the end of the sentence?

A faces
B see
C so
D happy

Assessment design

Measuring the construct

In developing items and designing the tests, the major criteria considered are as follows:

- validity of strands for measuring general ability
- distribution of items across strands
- distribution of item difficulty.

Distribution by strand

It is necessary to assess students on an appropriate distribution of strands so that the assessment provides a valid measure of general ability. This approach ensures that the formative data gained provides insight into possible strengths, gaps and weaknesses in different reasoning skills. The strands are evenly distributed across all levels to ensure that the general ability measures are not influenced by one strand over others. Table 1 shows the distribution of items by strand in each of the AGAT forms.

Table 1 *Number of AGAT items by strand for each test form*

Strands	Number of items Tests 1 – 2	Number of items Tests 3 – 9
Abstract reasoning	7	8
Kinetic reasoning	7	8
Numerical reasoning	7	8
Spatial reasoning	7	8
Verbal reasoning	7	8

Distribution of item difficulty

It is important to have a spread of item difficulties that match the abilities of the students. Table 2 shows the mean difficulty of the items in each of the AGAT forms in scale score units, with their standard deviations. Standard deviation measures the amount of variation in item difficulty for a set of items.

***Table 2** Mean difficulty and standard deviation of each AGAT form*

Test level	No. of items	Mean item difficulty (scale score)	Standard deviation (scale score)
1	35	111.6	8.2
2	35	115.2	7.7
3	40	117.8	9.2
4	40	120.7	9.9
5	40	121.8	8.3
6	40	124.3	9.1
7	40	125.5	8.1
8	40	127.6	10.4
9	40	129.1	9.5

Delivery

AGAT 2nd Edition is delivered solely online and can be administered any time throughout the year.

Choosing the right test

Planning and consistency are important in ensuring AGAT is used effectively and that students' results are useful and meaningful. The difficulty of a test and the teacher's knowledge of a student should be taken into consideration when selecting an appropriate test. There is often a wide range of ability within the classroom and it is not necessary to provide all students in a class with the same test. Instead, the focus should always be each student's ability at the time of the assessment, not where they are expected to be. Table 3 can be used as a guide to the appropriateness of a test for the year level being tested. Teachers should view the tests before assigning them to ensure appropriateness to the individual student.

Table 3 *Summary of test delivery details for AGAT*

Test level	Number of items*	Generally suitable for year levels	Time allowed**
1	35 (7 items per strand)	2, 3	50 minutes
2		2, 3, 4	
3	40 (8 items per strand)	3, 4, 5	
4		4, 5, 6	
5		5, 6, 7	
6		6, 7, 8	
7		7, 8, 9	
8		8, 9, 10	
9		9, 10, 11	

*All tests contain an equal number of items from each of the 5 strands – abstract reasoning, kinetic reasoning, numerical reasoning, spatial reasoning and verbal reasoning.

**The recommended time allocated for the test is 50 minutes, though some students may require additional time. In this case, an extra ten minutes can be given at the school's discretion.

Test items

Each test form comprises 35–40 items spread evenly across the five content strands. This means that there are seven or eight test items for each strand. This is just enough items to get a general inference for the student's ability in that cognitive strand and support the measurement of general ability. A strand-specific assessment should be undertaken for a more comprehensive study of a student's specific reasoning skills in a strand.

The items on each test form are arranged in an order to maximise student engagement, assessing numerous different skills from the different strands as the student progresses through the test. While most test items are individual stand-alone, some items that use the same stimulus appear next to each other to reduce the amount of reading required to orientate the student to the problem. Though each subsequent item may not statistically be more difficult than the one immediately preceding it, items generally increase in difficulty as the student moves through the test.

Frequency

Generally, AGAT is administered to students once per year. The AGAT scale enables teachers to monitor and compare students' general ability measures over time. For the purpose of monitoring student progress, a gap of 9 to 12 months between testing is recommended. General development may not be reflected in a student's AGAT scale scores over a shorter period of time. Longitudinal growth should be measured over a minimum of two years of schooling, or three separate testing sessions, in most contexts. This will help account for possible scale score variation, for example where external factors may affect a student's performance on a particular testing occasion.

Test administration

Teachers are required to supervise test administration. The recommended test administration time is 50 minutes. This should be sufficient for most students to complete their tests. If a student requires more time to complete the test, an additional 10 minutes can be given at the school's discretion. However, it is recommended that any additional time given is noted and taken into consideration when viewing reports.

Calculator use

None of the AGAT items should require the use of a calculator and it is recommended that students not be permitted to use calculators in any part of the test.

Item response formats

All items in AGAT 2nd Edition are multiple-choice format with four response options.

Reporting

The information provided by the AGAT reports is intended to assist teachers in understanding their students' general reasoning ability, strengths and weaknesses, and measuring general reasoning ability progress over time.

AGAT scale scores

An AGAT scale score is a numerical value given to a student whose ability has been measured by completing an AGAT assessment. A student's scale score lies at a point somewhere on the AGAT scale, and it indicates that student's level of ability – the higher the scale score, the higher the student's general ability.

Regardless of the test level or items administered to students, they will be placed on the same AGAT scale. This makes it possible to directly compare students' general abilities and to observe students' development by comparing their scale scores from multiple testing periods over time.

Item difficulty is a measure of the extent of skills and ability required to be successful on the item. This makes it possible to allocate each AGAT item a score on the same scale used to measure student ability. An item with a high scale score is more difficult for students to answer correctly than an item with a low scale score. It can generally be expected that a student is able to successfully respond to more items whose difficulty is located below their scale score than above. By referencing the difficulty of an item, or a group of items, and the proportion of correct responses by a student or within a group, it may be possible to identify particular items, or types of items, that have challenged students.

It is critical to note that the AGAT scale is not comparable to other scales such as PAT Maths or PAT Reading. For example, the AGAT scale score results for Test 3 **can** be compared with the scale scores from each of the other AGAT test levels. AGAT 1st Edition and AGAT 2nd Edition scale scores can also be compared, but neither edition of AGAT tests have any relation to PAT Maths Test scale score results but neither edition of AGAT tests have any relation to PAT Maths or PAT Reading scale score results.

AGAT scales are developed using the Rasch Measurement (RM) approach. See Appendix 1 for more information on the Rasch Measurement approach.

See Appendix 1 for more information on the Rasch Measurement approach.

General ability bands

While a scale score indicates a student's general ability according to the AGAT scale and can be used to quantitatively track a student's growth, it is only in understanding what the number represents that teachers can successfully inform their practice to support students. For this reason, the AGAT scale is divided into eight bands that demonstrate and describe general ability as a continuum.

The Table 4 displays the AGAT 2nd Edition general ability bands and the types of reasoning skills typically demonstrated at each band.

Table 4 *AGAT 2nd Edition general ability bands and descriptions*

General ability band	Scale score range	Band descriptions
Band 8	147 and above	Students typically can use deductive reasoning to solve multi-step numerical problems with different variables. They can use the relative dimensions of regular shapes and a knowledge of fractions to calculate the area of a shape expressed as a fraction of the whole. They can identify the missing output from a number machine involving complex quadratic expressions.
Band 7	139-147	Students typically can find the missing output from a number machine where two undisclosed calculations are applied. They can compare multiple different combinations of rules applied to intricate patterns in abstract settings. Students can identify the path in a complex network that results in a desired outcome. They can identify the intersecting points of multiple hidden shapes on a grid. Students can unpack complicated sentences and use sophisticated deductive reasoning to determine which statements are true and false.
Band 6	131-139	Students typically can solve numerical word-based problems involving unknown variables and multiple constraints, and can find the missing output in number machines involving simple quadratic relationships. They can identify the rules needed to transform one abstract pattern into another, and identify the missing step in a sequence of subtle images. Students can visualise the movement of an object through 3D space and locate objects on a grid after multiple non-routine commands. Students can identify a pair of words that are related in the same way as a given pair, and can compare uncommon words with very similar meanings. They can identify the rotations needed for two irregular shapes to tesselate, and can visualise how paper will look after a series of irregular folds and cuts.
Band 5	123-131	Students typically can solve numerical word-based problems involving the intersection of sets or multiple unknown variables. They can mentally transpose an object onto a grid to determine its location, and can identify how an object would appear from multiple perspectives. Students can use deductive reasoning in routine sentences to make comparisons between descriptions, and can distinguish between less familiar words with slightly different meanings. They can identify the next step in a sequence of complex moving images and apply different rules to a pattern to see how the pattern changes. Students can work backwards through a problem involving the movement of objects to determine the initial state. They can identify how cogs rotate in different scenarios.
Band 4	115-123	Students typically can solve simple image-based simultaneous equations, and can start to answer numerical word-based problems involving more than one unknown value. They can match abstract images that have similar features, and can identify the next step in a sequence of images with two rotating elements. Students can identify the turns needed for a vehicle to reach a point on a map, and can follow the path of an object through a system of tunnels. They can spot changes made to the arrangements of objects when photographed from different angles, and can recognise the order in which footprints were left in sand. Students can determine comparative attributes from simple sentences.

General ability band	Scale score range	Band descriptions
Band 3	107-115	Students typically can find the missing output in number machines involving simple linear relationships. They can identify missing numbers in a 3 x 3 table having been given the totals of rows or columns. They can recognise the order in which sheets of paper fell to the ground, and can identify the position of an object from a reverse perspective. Students can order a series of simple words based on subtleties in meaning and rearrange up to seven jumbled words to form a sentence. They can identify the starting point of an object in a grid after being given a set of movements and its end point. Students can predict the outcome of a change in mass distribution on a balance, and can identify the next step in a sequence of abstract patterns with a rotating element.
Band 2	99-107	Students typically can identify the missing number in a numerical sequence that increases by a small constant value. They can identify an overarching word that encompasses familiar given words, and can deduce the outcome of straightforward verbal comparisons. Students can spot simple patterns in abstract images and are starting to make connections between 2D and 3D pictures. They can identify the end point of an object in a grid after a given set of movements.
Band 1	99 and below	Students typically can identify the next number in a sequence of increasing even numbers. They can make connections between simple pairs of analogous words. Students can find the missing piece in a straightforward jigsaw puzzle. They can spot the picture needed to complete an abstract shape that involves simple reflection.

Norms [Comparison Groups]

AGAT norm data that represents the achievement of students across Australia is available as a reference against which student ability can be compared. The comparison between a student's scale score ability and the Australian norms is expressed as a percentile.

Appendices

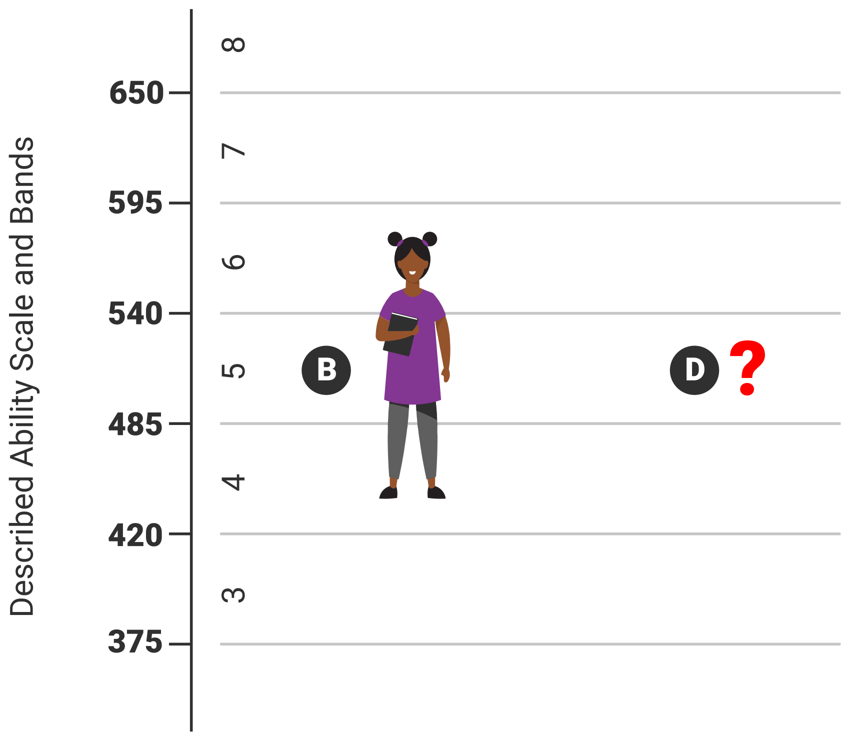
Appendix 1

The Rasch Measurement (RM) approach

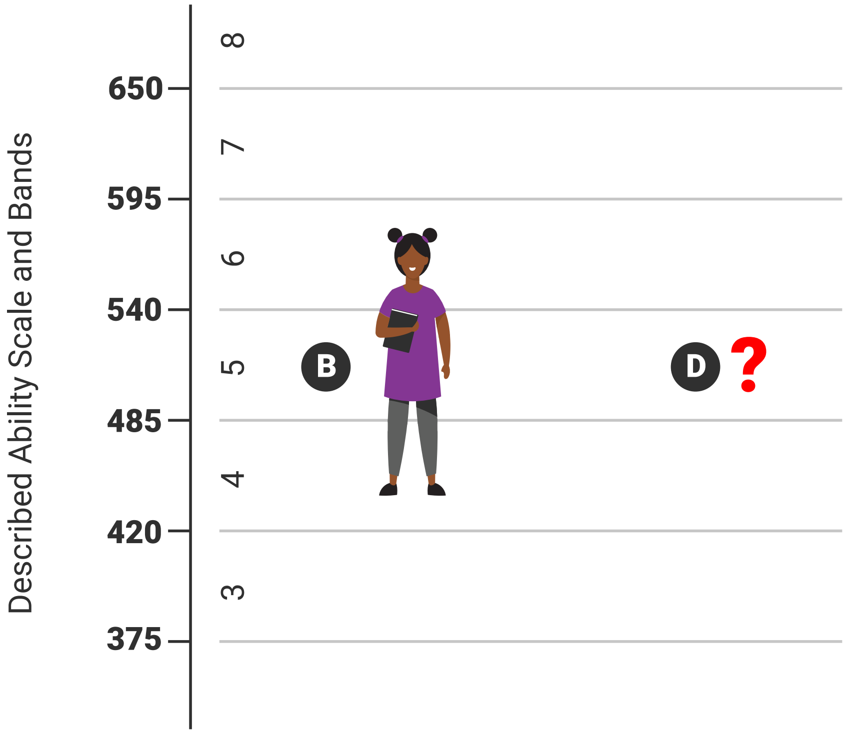
RM was developed by Danish mathematician Georg Rasch in the 1950s. It assumes that the ability of a student can be captured with a person parameter and the difficulty of an item (question) can be captured with an item parameter. Both parameters are reported as locations on the same scale because they measure the same construct: student ability reflects the level of skills and understandings demonstrated by the student; item difficulty reflects the level of skills and understandings required to answer the item correctly. The ability of a student is assumed to be completely represented by a location on the scale (person scale score); the difficulty of an item is also assumed to be completely represented by a location on the scale (item scale score). The location of a student on the scale is the same for all items and the location of an item is the same for all students. For example, an item answered correctly by 40 per cent of Grade 3 students is not necessarily more difficult than an item answered correctly by 70 per cent of Grade 9 students. The difference in item facility is due to the fact that the Grade 9 students are located higher on the scale, as it is assumed that they will have more experience with reasoning and problem solving both in the classroom context as well as in life, and not because the item has one location for the Grade 3 students and a different one for the Grade 9 students. The item has the same location on the scale for the Grade 3 students and the Grade 9 students, but the probability of success on this item is not the same for the two groups of students due to the different location of the Grade 3 and Grade 9 students on the scale.

RM proposes a functional dependence of the probability of success when a student located at a Point B on the scale attempts an item located at a Point D. The probability of success increases with student ability (location B) and it decreases with item difficulty (location D). It depends only on the student's and the item's locations on the scale, as follows:

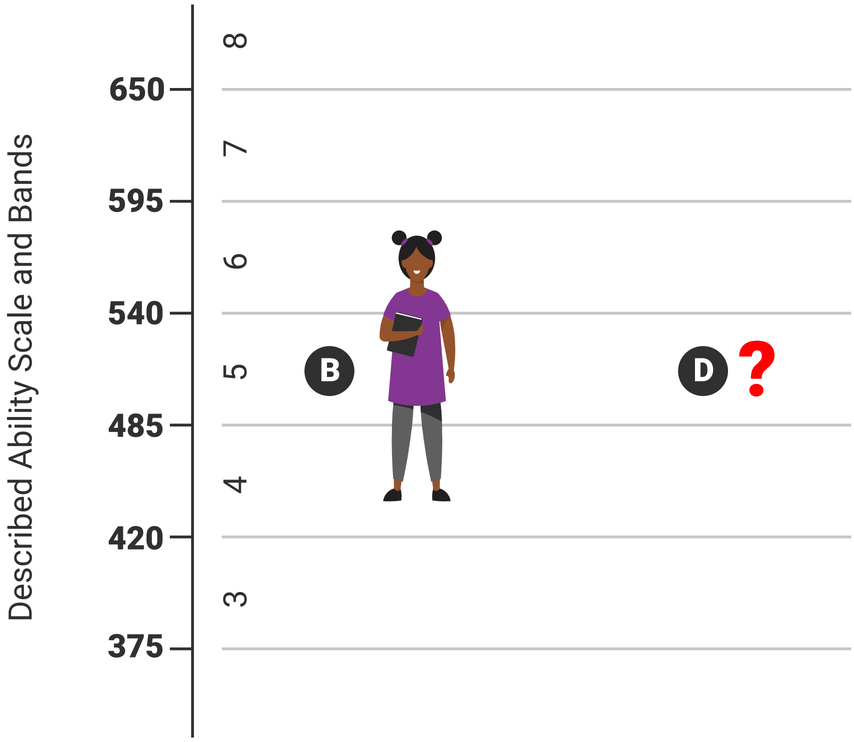
- If the student and the item are at the same location, it is equally likely for the student to answer it correctly or incorrectly (i.e., probability of success is 0.5)



- If the student is located higher on the scale than the item, it is likely for the student to answer the item correctly (i.e., probability is greater than 0.5)



- If the student is located lower on the scale than the item, it is not likely for the student to answer the item correctly (i.e., probability is less than 0.5). The greater the distance between the location of the student and the location of the item, the more the probability of success differs from 0.5.



Using RM, the AGAT measurement scale has been constructed on which student ability and item difficulty can be displayed. This scale represents a single construct of ability. All items used in the tests are used to describe levels of ability. Grade 3 students are located lower on the scale than students in higher year levels. Items in lower level tests, suitable for Grade 3 students, are located lower on the scale than items in difficult tests that are suitable for students in higher year levels.

The key features of the measurement scales are:

- Student ability and item difficulty are displayed on the same scale; the coordinates of locations on the scale (scale scores) are expressed numerically in the same unit; high-level ability and difficult items are located higher on the scale than lower levels of ability and easy items.
- Student ability is captured in raw scores, which are transformed into scale scores considering the difficulty of the test. Scale scores originating from scores on different tests (e.g., AGAT 2nd Edition Test 3 and Test 6) can be compared directly.
- The AGAT scales are interval scales: a change of one unit corresponds to the same amount of change in ability anywhere else on the scale. Raw scores and percentile ranks do not have this property: they are bunched closer together in the middle of their distributions than they are towards the ends.
- The location of items on the scale reflects their difficulty relative to other items and it is independent of the location of the distribution of ability of the students who provided the data. In other words, the distance between two items on the scale (their relative difficulty) is the same whether the data were collected from a highly able group or from a less able group.
- Levels of ability along the measurement scale can be qualitatively described, allowing the result of the assessment of a student to be reported in terms of qualitatively described levels of ability.
- The AGAT measurement scale has no boundaries, but most scale scores fall in the range 50 to 200.