PAT Vocabulary Skills

Technical report July 2019

Ling Tan
Eunjung Lee
Steven Kambouris
Clare Ozolins

Australian Council *for* Educational Research

Contents

Tables and figures

1 Overview

1.1 Background

The ACER Progressive Achievement Tests in Vocabulary Skills, commonly known as *PAT Vocabulary Skills*, are a set of online assessments of students' lexical knowledge. *PAT Vocabulary Skills* provides a measure of student achievement exclusively focused on vocabulary. The *PAT Vocabulary Skills* instrument has been constructed to reflect a broad conception of vocabulary that includes knowing, applying and categorising words, as well as understanding and using morphology (Anzai & Reinertsen, 2019). The test is designed to measure the lexical knowledge (including semantics and syntax) of Australian students from Years 2 to 10. Results are reported to teachers who use the data as an additional means to inform and measure the overall literacy abilities of students and plan for teaching and learning.

1.2 Trial test design

Nine trial test forms were developed and administered to students across Years 1 to 11. Test form lengths ranged from 31 to 40 items. The fact that the *PAT Vocabulary Skills* trial tests were delivered online allowed for the inclusion of interactive response formats. Each test form contained a mixture of multiple-choice items, interactive items and cloze items. Table 1 shows the total number of items in each trial test form and the number of items by item format. In each test form, the majority of items (56% to 79%) are multiple-choice items, including both simple multiple-choice questions and complex multiple-choice questions. The interactive items (drag-and-drop and hotspot items) account for 18% to 44% of items in a test. Only a small number of cloze items were trialled in test forms (up to 10% of items in a test).

Table 1 *Trial test forms and number of items by item type*

Trial test form	Multiple-choice	Drag-and-drop	Hotspot	Cloze	Total
PATV 2	20	9	2		31
PATV 23	19	14	1		34
PATV 34	22	11		2	35
PATV 45	26	9		4	39
PATV 56	24	13	2		39
PATV 67	25	13		2	40
PATV 78	31	7		1	39
PATV 89	30	7		1	38
PATV 910	24	12		4	40

1.3 Description of trial samples

The *PAT Vocabulary Skills* trial was conducted in 2017 with a sample of Australian schools. Test forms were administered online, delivered via ACER's Online Assessment and Reporting System (OARS). Schools using ACER's other online PAT assessments could opt for their students to participate in the test trial. The characteristics of the trial samples are detailed in Section 2 of this report.

1.4 Trial analysis

After the trial administration, student responses were analysed to assess the psychometric properties of all trial items and tests. During the initial analysis, the component responses in complex multiple-choice (CMC) items were split into separate responses and could then be treated as responses to separate items. This provided test developers an opportunity to diagnose the function of each part of the CMC items. After this initial analysis, responses to CMC items were scored as single items and analysed together with other non-CMC items.

Responses from each trial form were analysed separately using the Rasch model. These analyses indicated how well the items in each form fitted the Rasch measurement model and revealed items that did not perform as well as expected. A total of 245 *PAT Vocabulary Skills* items were trialled. Of these, 30 items (about 12%) were judged to have unsatisfactory psychometric properties and were deleted from the pool available for constructing the final test forms. The remaining items functioned well statistically and were well-targeted for difficulty as described by the test construct and assessment framework.

A common-item equating design was used to equate tests across year levels onto a single scale. This design made it possible to locate all items in the trial forms on a new scale, referred to as the *PAT Vocabulary Skills* scale. This means that student performance on different *PAT Vocabulary Skills* trial forms are directly comparable. Details of the equating design and equating results are provided in Section 7 of this report.

During *PAT Vocabulary Skills* test development, one goal was to avoid items that might favour one subgroup of students over another, for example girls compared with boys. Differential item functioning (DIF) analysis on gender was performed on all trial items. Any item exhibiting a statistically significant difference in subgroup performance for students of the same ability was flagged and subject to content analysis by test developers. Any items with content or context bias would potentially be excluded from the final *PAT Vocabulary Skills* assessment forms. Trial item analysis is described in more detail in Section 5 of this report.

1.5 Cut scores

The *PAT Vocabulary Skills* scale has been categorised into eight bands of achievement. Each band qualitatively describes the skills and understandings a student has demonstrated based on their performance on a *PAT Vocabulary Skills* test. The achievement bands are independent of the test forms. They can be used to compare student results obtained from different *PAT Vocabulary Skills* tests and assessed at different times. The determination of cut scores is described in the Section 8 of this report.

2 Description of trial samples

This section describes the demographic characteristics of the sample of students who participated the online trial in 2017. A total of 19 063 students from 171 schools across Australia participated the *PAT Vocabulary Skills* trial. Table 2 shows the number of students who participated in the trial by test form. The number of participants was smallest for trial form PATV910 (1345 students), and the highest in trial form PATV23 (2943 students).

Table 2 *Number of students by trial test form*

Year level	PATV 2	PATV 23	PATV 34	PATV 45	PATV 56	PATV 67	PATV 78	PATV 89	PATV 910	Total
Year 1	75	1								76
Year 2	1657	1685								3342
Year 3	1	1257	1291							2549
Year 4	1		1378	1471	4					2854
Year 5				1227	1310	1				2538
Year 6					1132	1130				2262
Year 7						911	963			1874
Year 8						13	788	778	23	1602
Year 9								644	690	1334
Year 10									612	612
Year 11									20	20
Total	1734	2943	2669	2698	2446	2055	1751	1422	1345	19 063

Overall, the proportion of female students (49%) and male students (51%) was similar. The proportion of male students was higher than female students by 5% or more at year levels 1 (57.9%), 7 (54.3%), and 10 (52.5%). Table 3 shows the proportion of students by gender at each year level.

Table 3 *Proportion of students in the trial sample by gender*

Year level	Total no. of students	Female (%)	Male (%)
Year 1	76	42.1	57.9
Year 2	3342	50.2	49.8
Year 3	2549	51.1	48.9
Year 4	2854	49.6	50.4
Year 5	2538	49.1	50.9
Year 6	2262	49.3	50.7
Year 7	1874	45.7	54.3
Year 8	1602	51.6	48.4
Year 9	1334	48.8	51.2
Year 10	612	47.5	52.5
Year 11	20	50.0	50.0
Total	19 063	49.4	50.6

Students from schools across Australia participated the *PAT Vocabulary Skills* trial. The majority of students (92%) were from Victoria (41%), Queensland (26%), and New South Wales (25%). The number of students from each state and territory is shown in Table 4.

Table 4 *Number of students in the trial sample by state and territory*

Year level	NSW	VIC	QLD	WA	SA	ACT	TAS	Total
Year 1	56	8	12	0	0	0	0	76
Year 2	673	1687	847	43	77	0	15	3342
Year 3	648	1207	507	70	96	0	21	2549
Year 4	812	1315	491	138	80	0	18	2854
Year 5	834	1063	458	56	83	22	22	2538
Year 6	663	956	344	155	71	24	49	2262
Year 7	301	705	633	141	73	21	0	1874
Year 8	283	472	768	57	0	0	22	1602
Year 9	250	329	689	66	0	0	0	1334
Year 10	275	44	154	122	0	0	17	612
Year 11	20	0	0	0	0	0	0	20
Total	4815	7786	4903	848	480	67	164	19 063

3 Data cleaning and data pre-processing

3.1 Data cleaning

Prior to data analysis, the item response data was checked for unexpected or invalid values. For example, valid codes of 'A', 'B', 'C' or 'D' were the expected responses to simple multiple-choice items. Response lengths and valid codes were also checked for hotspot items and drag-and-drop items. Item keys for simple multiple-choice items were checked for anomalies using item analysis statistics produced using ACER ConQuest software (Adams, Wu and Wilson, 2015). For example, for each item in a test form, the point biserial value for each option was checked to see if the correct response had the highest positive correlation with the total of the rest of item scores in the test.

3.2 Handling of missing data

Students may leave items unanswered either because an item was too difficult or because the student ran out of time and so did not attempt it. In the former case, the student is deemed to have seen the item and chosen not to provide a response. In the latter case, the student did not see the item at all. These two types of omitted or missing response data have been coded differently in the ACER online testing system.

If missing responses where students did not see the item are treated as incorrect responses, item difficulties may be overestimated. To avoid this, omitted responses on the items which were not seen by the students were treated as non-administered during item difficulty estimation. Both types of missing responses are considered as incorrect for the purpose of estimating student achievement scores.

3.3 Splitting of complex multiple-choice data

During the initial item analysis, the component responses of each complex multiple-choice (CMC) item were split into separate responses as if from multiple items. This analysis approach offers test developers opportunities to diagnose CMC item component performance and to determine the scoring rules for CMC items based on the empirical data.

3.4 Scoring of short response items

The frequencies of responses to short response items (cloze items) were tabulated and provided to test developers to review. This data was used to modify scoring rules if necessary. The updated scoring rules were then used for scoring the short response items for use in the item calibration data.

4 Scaling methodology

The response data from the *PAT Vocabulary Skills* trial tests was fitted to the partial credit Rasch measurement model (Rasch, 1980; Masters, 1982). This model is expressed mathematically as:

$$P(x_i|\theta_n) = \frac{\exp \sum_{j=0}^{x} (\theta_n - \delta_i + \tau_{ij})}{\sum_{h=0}^{m_i} \exp \sum_{j=0}^{h} (\theta_n - \delta_i + \tau_{ij})} \quad x_i = 0,1, \ldots, m_i$$

where $P(x_i|\theta_n)$ is the probability of person *n* to score *x* on item *i*. θ_n denotes the person's level of the latent trait, the item parameter δ_i gives the location of the item on the latent continuum, and τ_{ij} denotes *j*th step parameter of item *i*.

The Rasch model assumes that the achievement of a student can be captured with a person parameter and the difficulty of an item can be captured with an item parameter. The model allows both student achievement and item difficulty to be measured with the same scale. This is because they are both measurements of the same construct: student achievement reflects the level of skill and understanding demonstrated by the student; item difficulty reflects the level of skill and understanding required to answer the item correctly. High achieving students and difficult items are located higher on the scale than low achieving students and easy items.

PAT Vocabulary Skills scale scores can be used to directly compare a student's performance on two separate testing occasions, even if different test forms are used. This is possible because observed raw test scores on any *PAT Vocabulary Skills* test can be converted to locations on the *PAT Vocabulary Skills* scale. It is not meaningful to compare observed raw test scores from different test forms (even if they are expressed as percentages), because observed raw test scores and percentages do not take into account the relative difficulty of the tests.

The *PAT Vocabulary Skills* scale is an interval scale; a change of one unit corresponds to the same amount of change in achievement at all locations along the scale. The *PAT Vocabulary Skills* measurement scale has no upper or lower limits. When data is fitted to the Rasch model, the locations of items on the scale reflects their difficulty relative to other items and are independent of the distribution of student achievement along the scale. Levels of achievement along the measurement scale can be qualitatively described, allowing the result of the assessment of a student to be reported in descriptive terms.

The Rasch measurement modelling approach aims to have a test that collects data which fits the Rasch model. This is usually done during test development with vigorous test piloting and item selection processes. The Rasch model supports the construction of described achievement scales that not only report to students how well they are doing, but can also relate their performance to what they can typically do at their achievement level.

ACER ConQuest (Adams, Wu, and Wilson, 2015) was the software used for Rasch scaling analysis. This software provides tools for the estimation of a variety of different item response models and regression models. It was used for item calibration, and for generating weighted likelihood estimates (WLE) for person estimates. The transformations of student ability scores from logits to scale scores are presented in Appendix 2.

5 Trial item analysis

5.1 Item analysis

Initial analysis of the 245 unique items in the *PAT Vocabulary Skills* trial indicated that 30 items had poor fit to the model or inadequate discrimination. These items were removed from further analysis and consideration for the final test forms. Item statistics for the remaining trial items are provided in Appendix 1, including item difficulty (logit, scale score and facility), item discrimination (item rest correlation), item fit (weighted mean square, its confidence interval and T value) and the number of students attempting each item (number of data points).

A range of statistics are produced as part of the item analyses. Item facility and discrimination statistics are obtained from classical test analysis; the other statistics are obtained from item response theory analysis.

The *item facility* statistic expresses the percentage of individuals who are successful in answering each question or item on a test.

The *item discrimination* expresses the correlation between the individual item's score and the aggregate score on the set of items in the same test. The item discrimination index used throughout this report is item-rest correlation, in which the aggregate score excludes the score of the item under examination. The mean item-rest correlation for *PAT Vocabulary Skills* trial items was 0.43.

One of the *item difficulty* statistics is expressed in units of 'logits' – a metric used to measure the test results across different test forms on the same scale. Since individuals and items are measured on the same scale, judgements about the relative difficulty of the items can be made, along with judgements about the relative proficiency of students. Item difficulty can be expressed in terms of scale score points, and is achieved by transforming logits into scale score points as described in Appendix 2.

The *item fit* statistics measure the extent to which an item is contributing to the measurement of the characteristic of interest. In the case of the item weighted fit (referred to as weighted mean square), values near 1 are desirable. An item weighted fit value greater than one is often associated with a low discrimination index, and an item weighted fit value less than one is often associated with a high discrimination index. The mean of the weighted fit values for a given test calibration is 1.0.

The *item characteristic curve* visually depicts the relationship between probabilities of correct responses and differences between person ability and item difficulty. It can be shown that, when the observed item characteristic curve (ICC) is steeper than the expected ICC, the item fit mean-square value is less than one. When the observed ICC is flatter than the expected ICC, the item fit mean-square value is greater than one. The ICC of each *PAT Vocabulary Skills* item is available upon request to ACER.

The last column in the Appendix 1 table lists the number of students who saw the question. The minimum number of observations for any one item in the trial was 1204.

5.2 Differential item functioning

During item development, every effort is made to avoid producing items that might favour one subgroup of students over another. Despite this, a proportion of items may be flagged with potential differential item functioning (DIF) after statistical analysis. Investigating the reasons for a particular item showing DIF between particular groups involves looking for an explanatory connection between actual characteristics of the item and assumed or posited characteristics of the groups.

Gender DIF analysis was performed on all trial items by year level. The mean item difficulty in each of the two independent sets of item difficulties was centred at zero to adjust for group difference in ability. Any item in a subgroup with fewer than 100 observations was removed from DIF analysis, because of the small sample size.

Figure 1 shows the DIF plot for gender for each trial test form. On each DIF plot, an item is represented by one point on the plot. The red diagonal line serves as the reference line, with confidence interval limits indicated by the thin curved lines on either side of the reference line. If the relative item difficulty for an item is not different between the two groups (ie after taking their overall performance on the test into account), the point representing the item should lie on or close to the reference line. The distance of a point from the reference line indicates the magnitude of any potential DIF. Any item that falls outside the two lines representing the confidence interval limits may warrant investigation for potential DIF.

From Figure 1, it can be observed that the majority of items fall within the confidence interval limits or are close to the confidence interval limits. Only a few items are relatively far outside the confidence interval limits.

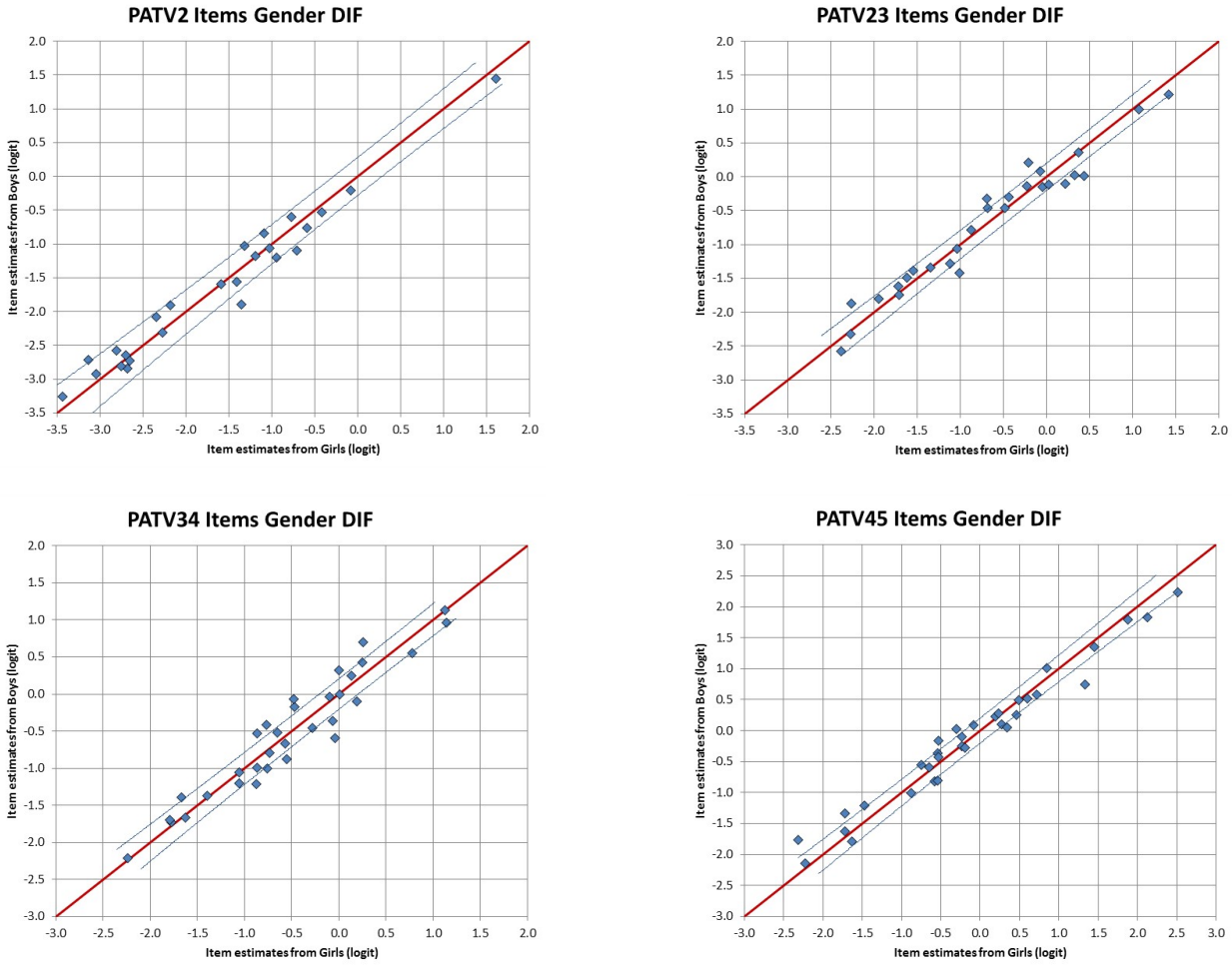
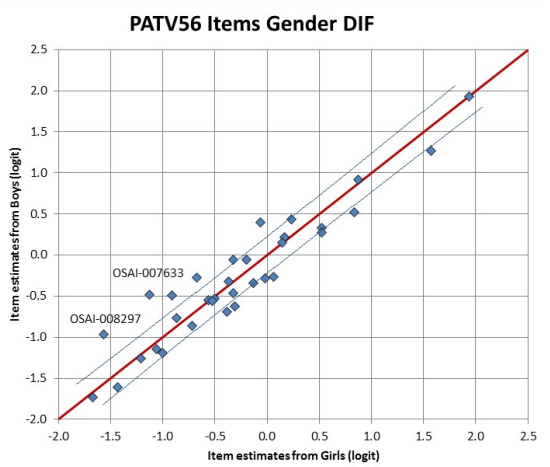
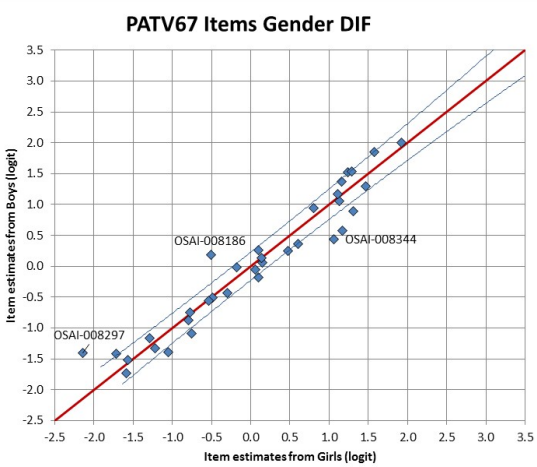
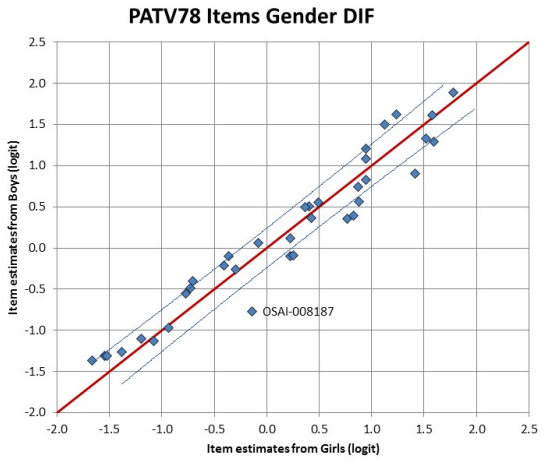
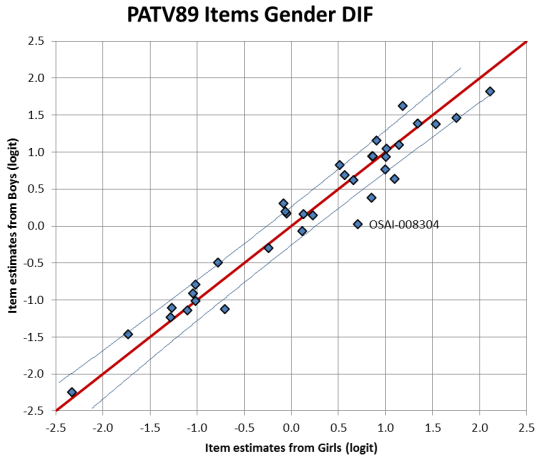


Figure 1 *Gender DIF plots for each trial test form (continued over)*









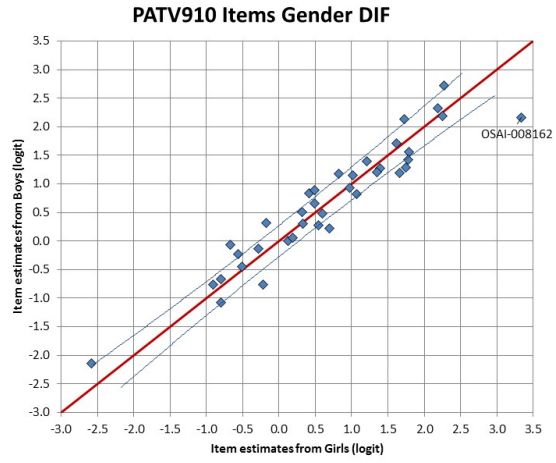


Figure 1 (continued) *Gender DIF plots for each trial test form*

Table 5 lists gender DIF analysis items with an absolute difference of 0.6 logits or larger and a standardised absolute difference of 5 or larger. These items were flagged for review. The difference for each item is calculated as the difficulty for the female students minus the difficulty for the male students. The table shows that four items statistically significantly favoured female students and four items statistically significantly favoured male students. Items showing DIF were investigated for biased content, and if bias were found to exist the items would not be selected for final tests. In practice, the DIF is often not content-related but rather performance-related; that is, the favoured subgroup is simply better at the skills being assessed, for a variety of reasons. After review, no trial items were removed for content bias.

Table 5 *List of potential gender DIF items*

Item label	Test form	Difference in item difficulties (logit)	Standardised difference in item difficulties	Chi-square	*p*-value	Gender favoured toward
OSAI-008297	PATV 56	-0.61	-5.07	25.71	0.00	female students
OSAI-007633	PATV 56	-0.65	-5.63	31.72	0.00	female students
OSAI-008297	PATV 67	-0.74	-5.34	28.56	0.00	female students
OSAI-008344	PATV 67	0.62	5.19	26.92	0.00	male students
OSAI-008186	PATV 67	-0.69	-5.80	33.60	0.00	female students
OSAI-008187	PATV 78	0.63	5.19	26.93	0.00	male students
OSAI-008304	PATV 89	0.68	5.15	26.48	0.00	male students
OSAI-008162	PATV 910	1.18	5.29	28.04	0.00	male students

6 Trial test analysis

6.1 Item–person maps

In addition to showing the locations of individual person abilities and item difficulties on the same scale, an item–person map provides a visual indication of test targeting. A test is targeted to the trial population if the test comprises items of varying difficulties, and the distribution of item difficulty is aligned to the distribution of student ability. On each item–person map, the distribution of students is plotted on the left side of the map, and the distribution of items is plotted on the right side. The higher-ability students and more difficult items are positioned towards the top of the scale, and the lower-ability students and easier items are positioned towards the bottom. Item–person maps for each test form are available upon request.

6.2 Test reliability

Test reliability indicates the extent to which a test is consistent in measuring what it is intended to measure (in this case, vocabulary skills). Test reliability does not imply validity, but it is a necessary condition for validity. The test reliability coefficient is equal to the proportion of observed raw score variance that is attributable to true scores. Two test reliability indices were calculated for the *PAT Vocabulary Skills* trial tests: Cronbach's alpha and expected a-posteriori/plausible value (EAP/PV) reliability, both shown in Table 6. The *PAT Vocabulary Skills* trial test reliabilities were 0.86 or higher in all trial tests.

Table 6 *Trial test reliabilities*

Test form	Cronbach's alpha coefficient	EAP/PV reliability
PATV 2	0.89	0.86
PATV 23	0.90	0.89
PATV 34	0.90	0.89
PATV 45	0.89	0.89
PATV 56	0.91	0.90
PATV 67	0.89	0.89
PATV 78	0.88	0.87
PATV 89	0.86	0.87
PATV 910	0.88	0.89

6.3 Correlations among strands

The *PAT Vocabulary Skills* assessment items can be categorised by the *process that they assess*. There are four process strands in *PAT Vocabulary Skills* that describe some key ways that students know and understand vocabulary. These four processes are Knowing, Applying, Categorising, and Morphology. Knowing describes knowing the meaning of words where there is no support that would enable a student to derive the meaning of the word from context. Applying describes recognising close relationships between different words such as identifying synonyms. Categorising describes the skill of sorting words into conceptual categories, or ordering words within a category by degree. Morphology describes the understanding and use of morphemes, sub-word level units of meaning such as affixes and roots, to support interpreting the meaning of words. Sometimes these affixes and roots are explicitly taught in schools, but the *PAT Vocabulary Skills* assessment has been developed on the principle that explicit teaching should not be necessary in order to arrive at the correct answer for these questions.

Table 7 shows the latent correlations between these process strands. The latent correlations do not have the problem of attenuation caused by measurement error in discrete ability estimates. The value of a correlation can range from –1.00 (perfect negative correlation) through 0.00 (no correlation) to 1.00 (perfect positive correlation). All the correlations shown in Table 7 are significant at the 0.01 level of confidence.

The correlations among *PAT Vocabulary Skills* strands across all year levels were estimated by fitting a multi-dimensional latent regression model using a Monte Carlo method in ConQuest. For each strand, delta-centred item difficulty parameters were estimated by fitting a unidimensional measurement model regressed on test level. The item difficulty estimates from all strands were then entered as anchored values in a multi-dimensional model regressed on test level.

Table 7 *Correlations between process strands and reliabilities of process strands*

	Categorising	Knowing	Applying	Morphology
Categorising	*.87*			
Knowing	.91	*.87*		
Applying	.89	.90	*.86*	
Morphology	.86	.85	.84	*.79*

Table 7 shows that there is a strong positive correlation (.84 or higher) between the process strands. The correlations among Knowing, Applying and Categorising (.89 or higher) are higher than the correlation between Morphology and other processes (.84 or higher), indicating that Morphology is a slightly different process from Knowing, Applying and Categorising. The italicised values in the table are the EAP/PV reliabilities for each process strand. The strong correlations among process strands indicate a coherent relationship among processes as defined by the assessment construct.

7 Trial equating design and results

The *PAT Vocabulary Skills* trial items were developed to target students from Year 2 to Year 10. Each *PAT Vocabulary Skills* item was trialled at two year levels, except for the items in the trial test form PATV2, which were targeted at students in Year 2 only.

About 25–30% of trial items served as common items for the purpose of equating tests across year levels onto the same scale. The process of equating test forms across different year levels is known as vertical equating. For *PAT Vocabulary Skills*, vertical equating was achieved through the placement of common items in test forms between adjacent year levels. Figure 2 shows the vertical equating design of the *PAT Vocabulary Skills* trial. Table 8 shows the number of vertical link items in each test form.

Table 8 *Number of vertical links in each trial test form*

Test form	Total number of items	Number of vertical link items – link to higher level
PATV 2	31	10
PATV 23	34	10
PATV 34	35	10
PATV 45	39	10
PATV 56	39	10
PATV 67	40	11
PATV 78	39	10
PATV 89	38	10
PATV 910	40	NA
Total	346	

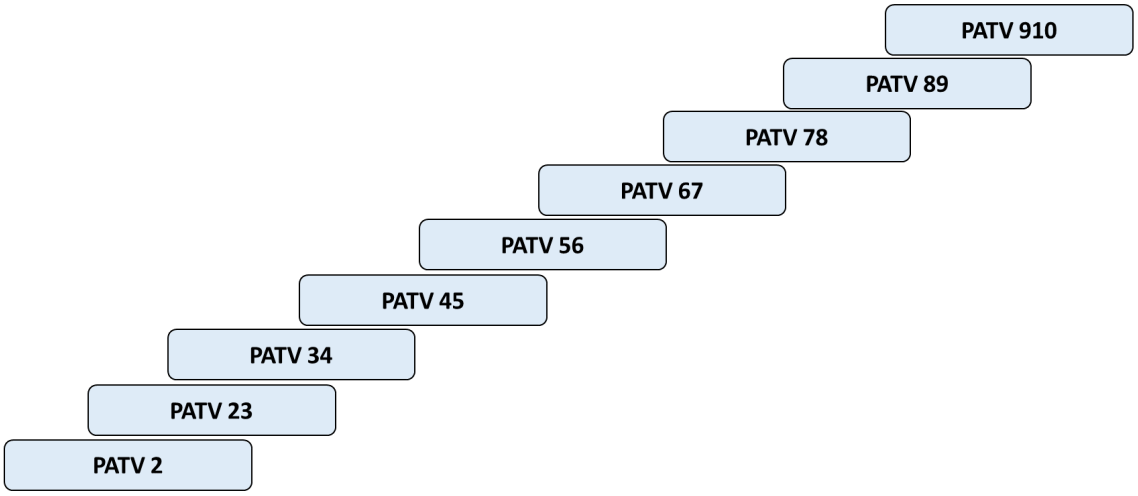


Figure 2 *Vertical equating design for the PAT Vocabulary Skills trial*

Common items between adjacent year levels were examined for the ordering of relative item difficulties in both year levels. This was to check whether common items between adjacent year levels were working as intended, and to confirm the validity of the vertical equating.

Figure 3 to Figure 10 show the scatter plots examining vertical equating by plotting the relative difficulties of common items between adjacents *PAT Vocabulary Skills* test forms. In each figure, the left panel shows the results before reviewing common items, and the right panel shows the results after excluding any misfitting items and any items with a standardised difference greater than 3. The standardised difference is the difference of item difficulty estimates (adjusted for year-level differences) divided by the pooled standard error. In each plot, the mean item difficulty in each of two sets of item difficulties was set to be the same to adjust for year-level differences in ability. It can be observed that the vertically linked items were scattered around the diagonal identity line. The vertically linked items in each chart covered a wide range of item locations spanning at least 1.5 logits. The plots indicated that *PAT Vocabulary Skills* vertically-linked common items were working well.

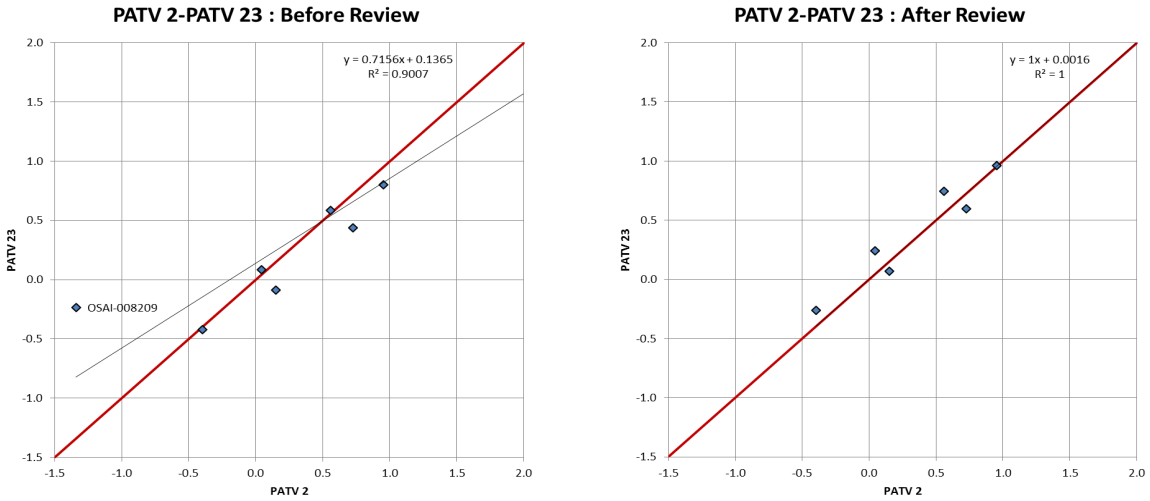


Figure 3 *Vertical equating link item review for PAT Vocabulary Skills trial form 2 and form 23*

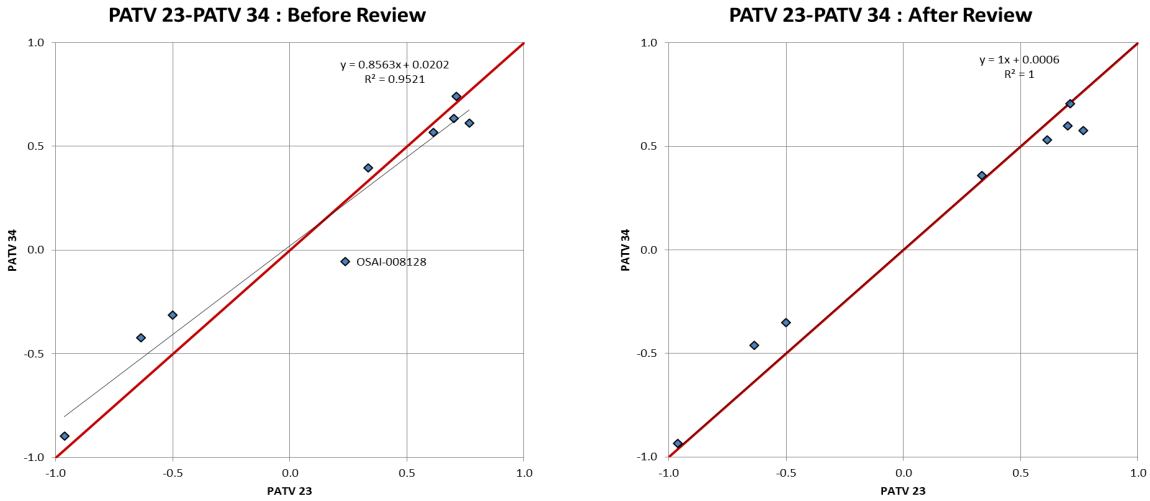


Figure 4 *Vertical equating link item review for PAT Vocabulary Skills trial form 23 and form 34*

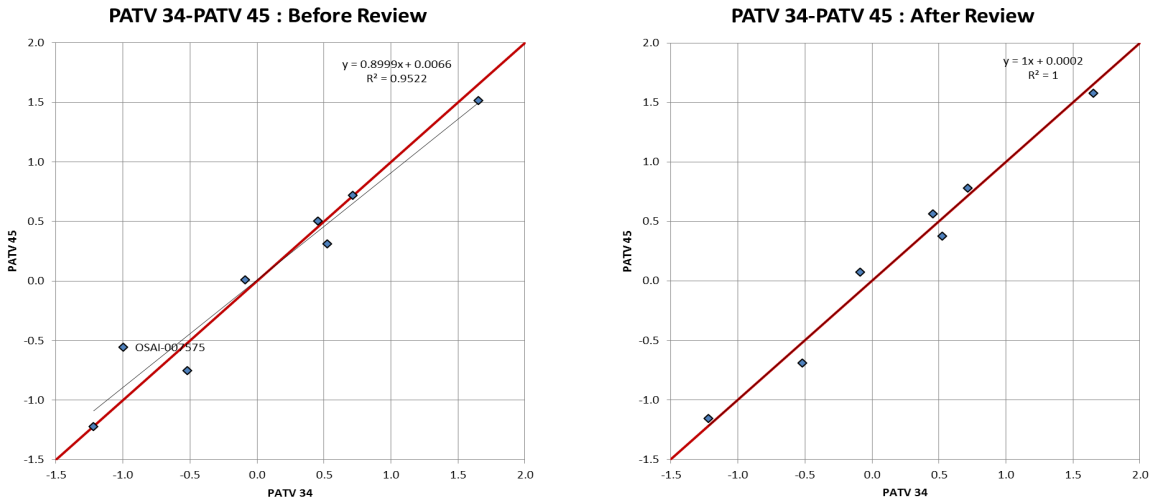


Figure 5 *Vertical equating link item review for PAT Vocabulary Skills trial form 34 and form 45*

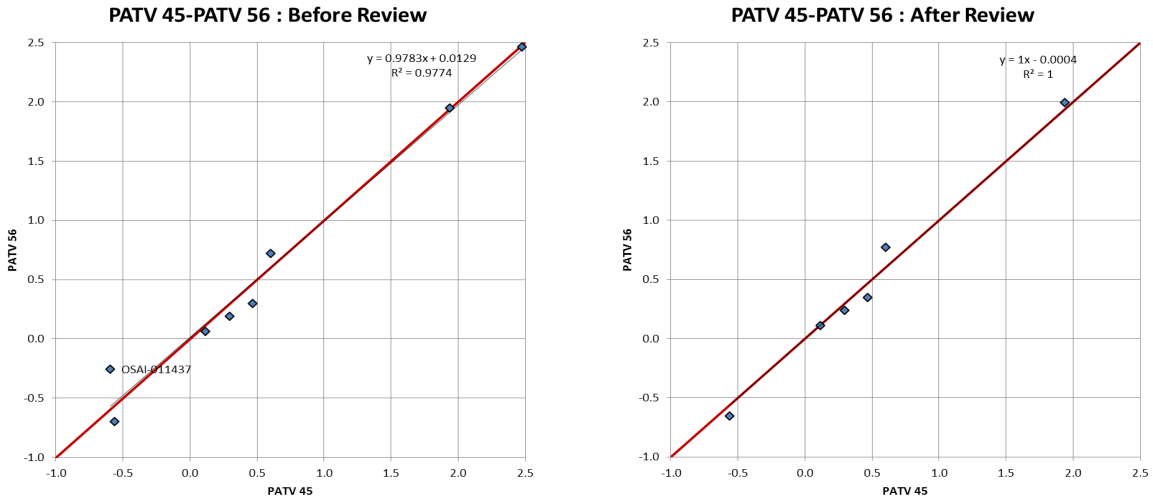


Figure 6 *Vertical equating link item review for PAT Vocabulary Skills trial form 45 and form 56*

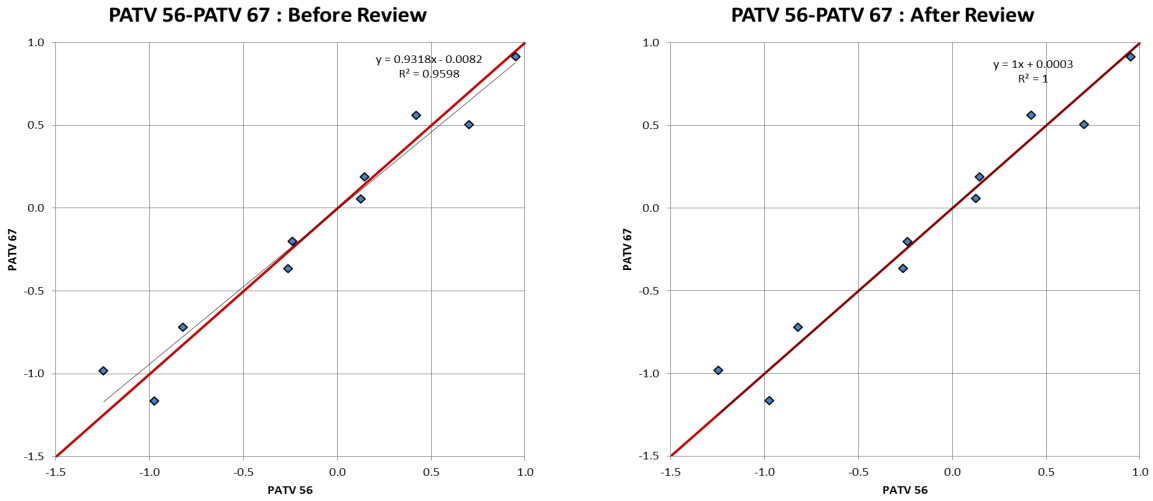


Figure 7 *Vertical equating link item review for PAT Vocabulary Skills trial form 56 and form 67*

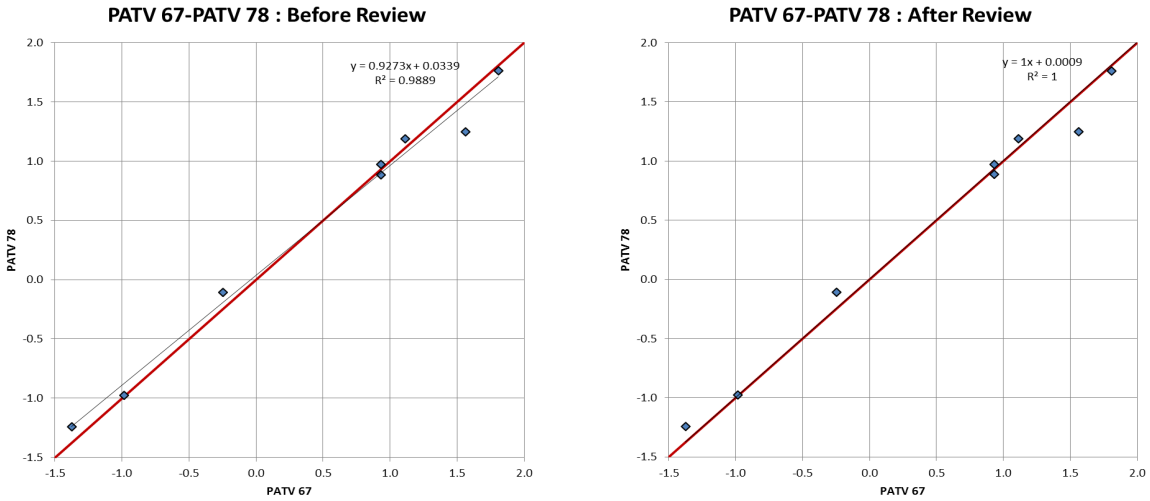


Figure 8 *Vertical equating link item review for PAT Vocabulary Skills trial form 67 and form 78*

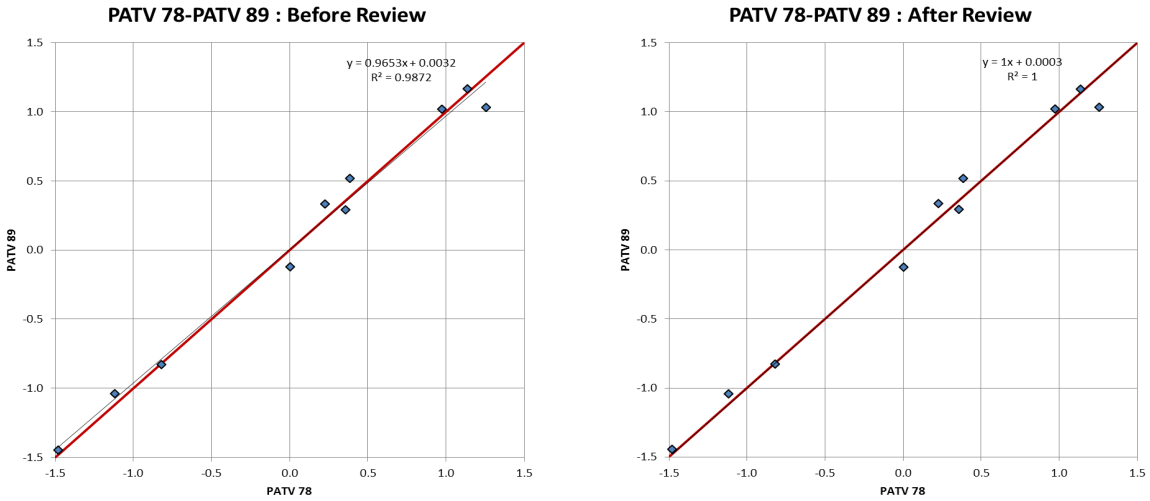


Figure 9 *Vertical equating link item review for PAT Vocabulary Skills trial form 78 and form 89*

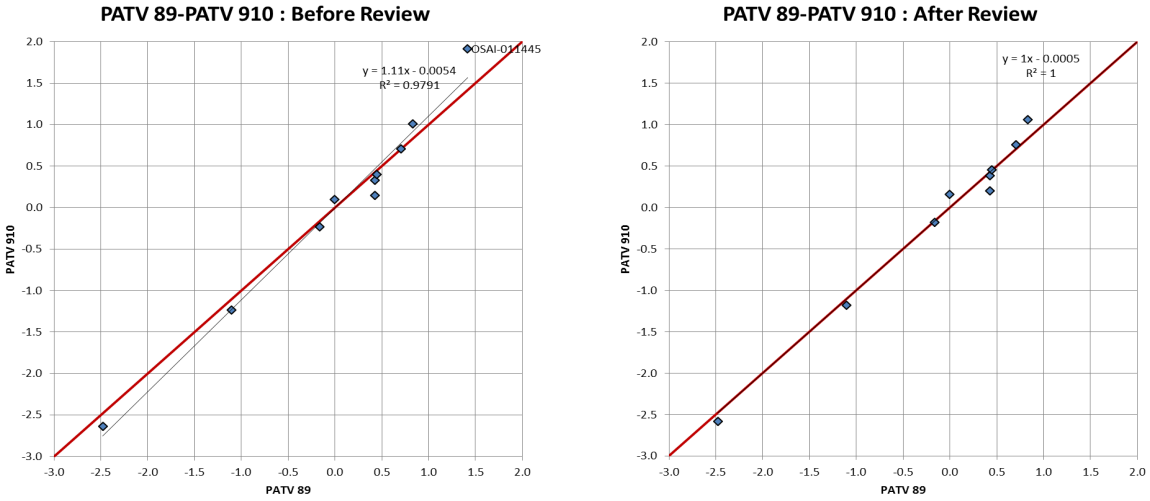


Figure 10 *Vertical equating link item review for PAT Vocabulary Skills trial form 89 and form 910*

8 Setting cut scores for achievement bands

Eight achievement bands were established for the *PAT Vocabulary Skills* assessments, based on the *PAT Vocabulary Skills* scale. Each band has a width of 10 scale score points. The cut scores defining the thresholds between adjacent bands are shown in Table 9. The lower cut score is inclusive. The percentages of *PAT Vocabulary Skills* trial participants located within in each band are shown in Table 10, broken down by year level. The *PAT Vocabulary Skills* achievement bands are described in detail by Anzai & Reinertsen (2019).

Table 9 *PAT Vocabulary Skills achievement band cut scores*

Bands	Lower cut (scale score)	Upper cut (scale score)
Band 8	≥150	
Band 7	140	150
Band 6	130	140
Band 5	120	130
Band 4	110	120
Band 3	100	110
Band 2	90	100
Band 1		<90

Table 10 *Percentages of trial participants within each achievement band by year level*

Band	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8	Year 9	Year 10
Band 8			1	2	2	6	8	14
Band 7		2	4	10	12	20	23	32
Band 6	2	8	15	27	31	35	34	35
Band 5	13	23	31	34	35	27	25	16
Band 4	30	32	30	21	17	10	9	3
Band 3	33	24	15	6	3	2	1	
Band 2	17	9	4	1				
Band 1	4	2						

9 PAT Vocabulary Skills test forms

Following the detailed trial analysis, five *PAT Vocabulary Skills* test forms were constructed. Each form begins with practice items, followed by assessment items. It is anticipated that completing one test form (including both the practice items and assessment items) will take approximately one hour.

The practice items have been designed to introduce test takers to the online testing interface and the item formats that appear in each test form. The practice items are intentionally easy, so that students are not distracted by the content. They allow students to practise responding to multiple-choice items presented in a variety of formats, multiple-choice items presented as rows in tables that require a response for each row, dragging and dropping objects, selecting a hotspot, and scrolling down the page to see more content. Once students have completed the practice items, they are able to move on to complete the assessment items.

From the Rasch measurement analyses that were carried out, it is possible to report the mean difficulty of the items in each of the *PAT Vocabulary Skills* tests in scale score units. These mean item difficulties – or test difficulties – are shown together with their standard deviations in Table 11.

Table 11 *PAT Vocabulary Skills test difficulties*

Test form	No. of items	Mean item difficulty (scale score)	Standard deviation (scale score)
Test 1	27	93.3	9.5
Test 2	30	106.2	8.3
Test 3	33	118.3	7.8
Test 4	35	130.4	8.0
Test 5	38	140.4	7.6

The locations of the *PAT Vocabulary Skills* items on the measurement scale are shown in the separate document *PAT Vocabulary Skills Test and Item Difficulty*. Overlap in the difficulty of items and processes, as well as an overall progression in difficulty over the course of the tests, can be clearly seen.

Score equivalence tables for *PAT Vocabulary Skills* tests were created based on the *PAT Vocabulary Skills* trial item parameters estimated from concurrent calibration. After checking vertical links, response data from different year levels can be combined into a single data file for the concurrent calibration. The concurrent calibration places all trial items from different year levels on the same scale simultaneously in a single calibration. ConQuest was used to create tables showing the equivalence between raw scores and ability estimates expressed in logit values. The ability estimates were next transformed to the *PAT Vocabulary Skills* scale using logit-to-scale score transformation parameters. The scale score transformation formula is shown in Appendix 2. The score equivalence tables for the *PAT Vocabulary Skills* tests are shown in Appendix 3.

References

Adams, R.J; Wu, M.L & Wilson, M.R. (2015). *ConQuest 4*. [computer program] Camberwell: Australian Council for Educational Research.

Masters, G.N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149–174.

Anzai, D. & Reinertsen, N. (2019). *PAT Vocabulary Skills* Domain Outlines, Australian Council for Educational Research.

Rasch, G (1980). *Probabilistic Models for Some Intelligence and Attainment Tests*. Chicago: MESA Press (original work published 1960).

Appendices

Appendix 1

PAT Vocabulary Skills trial item statistics

Item label	Process	Item difficulty estimate (logit)	Item difficulty estimate (scale score)	Facility	Discrimination (Item–Rest corr.)	Weighted Fit		No. of data points
						Weighted mean square (95% confidence interval)	T value	
OSAI-008130	Categorising	-3.620	83.8	80.8	0.51	0.96 (0.92, 1.08)	-1.1	1724
OSAI-008190	Knowing	-4.080	79.2	85.6	0.44	1.01 (0.91, 1.09)	0.2	1724
OSAI-011387	Knowing	-4.705	72.9	90.7	0.55	0.82 (0.88, 1.12)	-3.1	1724
OSAI-008125	Knowing	-0.313	116.9	28.4	0.35	0.98 (0.97, 1.03)	-0.9	4642
OSAI-008213	Knowing	-2.539	94.6	67.2	0.49	0.95 (0.96, 1.04)	-2.8	4642
OSAI-008209_2	Knowing	-4.363	76.4	88.1	0.43	0.97 (0.90, 1.10)	-0.6	1724
OSAI-008209_23	Knowing	-3.292	87.1	78.9	0.48	0.97 (0.94, 1.06)	-1.0	2918
OSAI-007564	Applying	-3.290	87.1	75.1	0.60	0.88 (0.96, 1.04)	-5.4	4642
OSAI-007560	Knowing	-4.000	80.0	81.8	0.62	0.96 (0.91, 1.09)	-0.8	1724
OSAI-007562	Applying	-3.500	85.0	85.5	0.60	0.65 (0.92, 1.08)	10.1	1724
OSAI-007569	Applying	-2.470	95.3	64.0	0.39	1.15 (0.94, 1.06)	4.8	1724
OSAI-008123	Categorising	-3.074	89.3	73.8	0.53	0.95 (0.93, 1.07)	-1.5	1724
OSAI-008291	Knowing	-2.652	93.5	67.5	0.49	1.01 (0.94, 1.06)	0.4	1724
OSAI-008195	Categorising	-2.511	94.9	66.8	0.47	1.01 (0.96, 1.04)	0.7	4642
OSAI-011380	Knowing	-3.002	90.0	74.2	0.52	0.93 (0.96, 1.04)	-3.7	4642
OSAI-008210	Knowing	-2.239	97.6	64.1	0.33	1.13 (0.97, 1.03)	7.6	4459
OSAI-011382	Applying	-3.469	85.3	80.3	0.56	0.83 (0.95, 1.05)	-7.4	4642
OSAI-011383	Knowing	-2.522	94.8	65.4	0.52	0.97 (0.94, 1.06)	-1.0	1724
OSAI-008296	Knowing	-4.287	77.1	87.5	0.51	0.90 (0.90, 1.10)	-2.0	1724
OSAI-011378	Applying	-4.092	79.1	85.7	0.58	0.83 (0.91, 1.09)	-3.8	1724
OSAI-008193	Knowing	-4.161	78.4	86.4	0.60	0.79 (0.90, 1.10)	-4.5	1724
OSAI-007573	Applying	-2.060	99.4	67.2	0.44	1.03 (0.95, 1.05)	1.0	1724
OSAI-007561	Knowing	-3.830	81.7	86.5	0.59	0.69 (0.91, 1.09)	-8.2	1724
OSAI-007583	Knowing	-2.177	98.2	59.6	0.42	1.10 (0.95, 1.05)	3.6	1724
OSAI-007568	Knowing	-1.663	103.4	50.4	0.44	1.00 (0.95, 1.05)	0.1	1724
OSAI-007570	Knowing	-1.976	100.2	56.0	0.55	0.89 (0.95, 1.05)	-4.6	1724
OSAI-008153	Morphology	-1.552	104.5	56.5	0.47	0.99 (0.97, 1.03)	-1.0	5570
OSAI-008150	Knowing	-3.066	89.3	80.0	0.43	1.00 (0.96, 1.04)	0.1	5570
OSAI-008113	Knowing	-1.418	105.8	54.1	0.46	1.01 (0.97, 1.03)	0.6	5570
OSAI-008211	Categorising	-2.676	93.2	74.9	0.44	1.03 (0.96, 1.04)	1.4	5570
OSAI-011371	Knowing	-1.779	102.2	60.5	0.47	1.00 (0.97, 1.03)	0.1	5570
OSAI-008135	Categorising	-1.451	105.5	54.7	0.42	1.04 (0.97, 1.03)	2.8	5570
OSAI-011399	Morphology	-2.554	94.5	73.1	0.34	1.15 (0.96, 1.04)	8.0	5570
OSAI-008356	Knowing	-1.475	105.2	55.1	0.45	1.01 (0.97, 1.03)	0.9	5570
OSAI-008128_23	Applying	-1.920	100.8	57.6	0.54	0.94 (0.96, 1.04)	-3.2	2918
OSAI-008128_34	Applying	-2.179	98.2	72.9	0.59	0.82 (0.95, 1.05)	-7.5	2652

Item label	Process	Item difficulty estimate (logit)	Item difficulty estimate (scale score)	Facility	Discrimination (Item–Rest corr.)	Weighted Fit		No. of data points
						Weighted mean square (95% confidence interval)	T value	
OSAI-007565	Applying	-3.711	82.9	83.7	0.56	0.83 (0.93, 1.07)	-5.3	2918
OSAI-008214	Knowing	-1.495	105.0	49.9	0.46	1.02 (0.96, 1.04)	1.1	2918
OSAI-008131	Categorising	-1.641	103.6	52.6	0.46	1.03 (0.96, 1.04)	1.5	2918
OSAI-008358	Categorising	-1.277	107.2	45.9	0.45	1.05 (0.96, 1.04)	2.6	2918
OSAI-008124	Categorising	-0.155	118.5	26.7	0.39	1.02 (0.95, 1.05)	0.7	2918
OSAI-011397	Knowing	-2.022	99.8	59.5	0.48	1.02 (0.96, 1.04)	0.8	2918
OSAI-008192	Knowing	-3.885	81.1	85.5	0.54	0.83 (0.93, 1.07)	-4.9	2918
OSAI-008293	Applying	-2.895	91.1	73.6	0.56	0.88 (0.95, 1.05)	-4.7	2918
OSAI-007584	Knowing	-1.100	109.0	42.7	0.50	0.96 (0.96, 1.04)	-2.1	2918
OSAI-007571	Applying	-2.210	97.9	58.4	0.44	1.12 (0.96, 1.04)	5.7	2918
OSAI-007581	Applying	-1.400	106.0	45.0	0.35	1.16 (0.96, 1.04)	8.3	2918
OSAI-007575	Applying	-1.900	101.0	75.0	0.34	1.02 (0.95, 1.05)	1.0	2652
OSAI-007590	Knowing	0.400	124.0	36.1	0.27	1.36 (0.95, 1.05)	13.8	2652
OSAI-008202	Applying	0.313	123.1	34.4	0.29	1.09 (0.97, 1.03)	6.3	5169
OSAI-008127	Knowing	-1.319	106.8	63.5	0.44	1.01 (0.97, 1.03)	0.7	5322
OSAI-008164	Morphology	-1.906	100.9	75.2	0.42	0.93 (0.96, 1.04)	-3.8	5098
OSAI-008368	Morphology	-0.795	112.1	53.9	0.50	0.94 (0.97, 1.03)	-4.6	5322
OSAI-008315	Applying	-0.556	114.4	49.4	0.32	1.14 (0.97, 1.03)	10.1	5322
OSAI-008163	Categorising	-2.503	95.0	81.1	0.53	0.80 (0.96, 1.04)	-9.3	5322
OSAI-008386	Applying	-0.854	111.5	55.0	0.55	0.85 (0.97, 1.03)	11.4	5322
OSAI-008126	Categorising	0.266	122.7	29.3	0.27	1.18 (0.96, 1.04)	7.6	2652
OSAI-008288	Applying	-1.163	108.4	55.5	0.54	0.92 (0.96, 1.04)	-4.1	2652
OSAI-008152	Knowing	-1.674	103.3	64.7	0.54	0.91 (0.96, 1.04)	-4.1	2652
OSAI-008166	Knowing	-1.850	101.5	67.7	0.49	0.96 (0.96, 1.04)	-1.6	2652
OSAI-008370	Applying	-1.007	109.9	52.5	0.48	0.98 (0.96, 1.04)	-1.0	2652
OSAI-008145	Knowing	-1.107	108.9	54.4	0.39	1.09 (0.96, 1.04)	4.8	2652
OSAI-008357	Knowing	-0.456	115.4	42.1	0.46	1.00 (0.96, 1.04)	0.2	2652
OSAI-008114	Knowing	-1.117	108.8	54.6	0.57	0.88 (0.96, 1.04)	-6.5	2652
OSAI-008151	Categorising	-1.065	109.4	53.6	0.56	0.89 (0.96, 1.04)	-6.2	2652
OSAI-008144	Categorising	-0.320	116.8	39.6	0.36	1.11 (0.96, 1.04)	5.4	2652
OSAI-007566	Applying	-1.925	100.7	68.9	0.42	1.04 (0.95, 1.05)	1.6	2652
OSAI-007582	Applying	-1.300	107.0	45.5	0.35	1.25 (0.96, 1.04)	11.9	2652
OSAI-007586	Applying	-0.600	114.0	47.5	0.52	0.93 (0.96, 1.04)	-3.9	2652
OSAI-007579	Knowing	-1.600	104.0	59.5	0.55	0.95 (0.96, 1.04)	-2.5	2652
OSAI-011415	Categorising	0.270	122.7	43.1	0.35	1.14 (0.97, 1.03)	9.5	5090
OSAI-011423	Knowing	2.077	140.8	15.7	0.16	1.14 (0.95, 1.05)	5.2	5090
OSAI-011433	Categorising	-0.141	118.6	50.8	0.41	1.07 (0.97, 1.03)	4.7	5090
OSAI-011436	Applying	0.001	120.0	48.1	0.49	0.97 (0.97, 1.03)	-2.5	5090
OSAI-011437_45	Categorising	-0.981	110.2	62.7	0.37	1.10 (0.96, 1.04)	4.9	2670
OSAI-011437_56	Categorising	-0.643	113.6	64.0	0.34	1.16 (0.96, 1.04)	7.0	2420
OSAI-011438	Categorising	-0.297	117.0	53.7	0.43	1.05 (0.97, 1.03)	3.9	5090
OSAI-011439	Categorising	1.552	135.5	22.1	0.19	1.23 (0.96, 1.04)	10.7	5090
OSAI-011440	Categorising	-1.008	109.9	66.4	0.51	0.93 (0.97, 1.03)	-4.4	5090
OSAI-008365	Categorising	-0.080	119.2	45.8	0.54	0.90 (0.96, 1.04)	-5.7	2670
OSAI-008314	Applying	-0.509	114.9	53.9	0.52	0.94 (0.96, 1.04)	-3.4	2670
OSAI-008366	Categorising	-0.614	113.9	55.9	0.42	1.05 (0.96, 1.04)	2.8	2670
OSAI-008189	Knowing	0.380	123.8	39.1	0.37	1.06 (0.96, 1.04)	3.1	2520

Item label	Process	Item difficulty estimate (logit)	Item difficulty estimate (scale score)	Facility	Discrimination (Item–Rest corr.)	Weighted Fit		No. of data points
						Weighted mean square (95% confidence interval)	T value	
OSAI-008180	Applying	-0.754	112.5	61.2	0.55	0.87 (0.96, 1.04)	-7.1	2497
OSAI-011398	Categorising	-1.951	100.5	78.1	0.57	0.81 (0.94, 1.06)	-7.2	2670
OSAI-008359	Morphology	-0.896	111.0	61.1	0.45	1.02 (0.96, 1.04)	0.9	2670
OSAI-008345	Knowing	-0.060	119.4	45.4	0.46	0.99 (0.96, 1.04)	-0.4	2670
OSAI-011441	Categorising	1.707	137.1	17.1	0.20	1.13 (0.94, 1.06)	3.7	2670
OSAI-011443	Categorising	1.132	131.3	24.8	0.33	1.06 (0.95, 1.05)	2.3	2670
OSAI-011444	Categorising	0.659	126.6	32.3	0.40	1.02 (0.96, 1.04)	0.7	2670
OSAI-007620	Applying	0.510	125.1	44.5	0.51	1.06 (0.96, 1.04)	2.9	2670
OSAI-007653	Knowing	0.730	127.3	30.7	0.30	1.13 (0.96, 1.04)	6.0	2670
OSAI-007635	Applying	-0.780	112.2	52.0	0.51	1.02 (0.96, 1.04)	0.9	2670
OSAI-007644	Knowing	-0.210	117.9	52.7	0.57	0.89 (0.96, 1.04)	-6.5	2670
OSAI-008367	Morphology	0.351	123.5	48.1	0.47	0.99 (0.97, 1.03)	-0.9	4456
OSAI-008297	Knowing	-1.184	108.2	74.9	0.39	1.00 (0.96, 1.04)	-0.1	4456
OSAI-008116	Categorising	-0.039	119.6	55.5	0.48	0.97 (0.97, 1.03)	-2.4	4456
OSAI-008305	Categorising	-0.354	116.5	61.2	0.51	0.93 (0.97, 1.03)	-4.6	4456
OSAI-008196	Categorising	0.797	128.0	39.9	0.39	1.07 (0.97, 1.03)	4.5	4456
OSAI-008382	Categorising	-1.255	107.4	76.0	0.41	0.98 (0.96, 1.04)	-1.1	4456
OSAI-008362	Morphology	0.474	124.7	45.9	0.43	1.02 (0.97, 1.03)	1.3	4456
OSAI-008363	Morphology	0.034	120.3	54.1	0.46	1.00 (0.97, 1.03)	0.0	4456
OSAI-007605	Applying	-0.906	110.9	70.7	0.47	0.95 (0.96, 1.04)	-3.1	4456
OSAI-007642	Knowing	-0.430	115.7	62.8	0.52	0.89 (0.97, 1.03)	-7.3	4456
OSAI-007633	Knowing	-0.648	113.5	64.1	0.42	1.05 (0.96, 1.04)	2.3	2420
OSAI-007613	Applying	-0.700	113.0	59.9	0.55	0.97 (0.96, 1.04)	-1.4	2420
OSAI-008200	Categorising	0.461	124.6	43.8	0.50	0.97 (0.96, 1.04)	-1.7	2420
OSAI-011446	Categorising	-1.541	104.6	78.0	0.38	1.02 (0.94, 1.06)	0.7	2420
OSAI-008316	Categorising	0.039	120.4	51.6	0.43	1.07 (0.96, 1.04)	3.3	2420
OSAI-011448	Applying	-0.255	117.5	57.0	0.64	0.81 (0.96, 1.04)	10.1	2420
OSAI-011457	Morphology	-0.957	110.4	69.3	0.52	0.93 (0.95, 1.05)	-3.2	2420
OSAI-008198	Knowing	-0.556	114.4	62.4	0.54	0.92 (0.96, 1.04)	-3.9	2420
OSAI-011449	Categorising	1.025	130.2	33.8	0.47	0.99 (0.96, 1.04)	-0.4	2420
OSAI-008194	Categorising	0.297	123.0	46.8	0.51	0.97 (0.96, 1.04)	-1.3	2420
OSAI-007612	Knowing	-0.398	116.0	59.6	0.56	0.90 (0.96, 1.04)	-5.2	2420
OSAI-007618	Applying	0.000	120.0	53.4	0.47	1.01 (0.96, 1.04)	0.5	2420
OSAI-007614	Applying	-0.700	113.0	64.5	0.57	0.88 (0.96, 1.04)	-5.5	2420
OSAI-007645	Applying	-0.325	116.7	58.3	0.54	0.93 (0.96, 1.04)	-3.6	2420
OSAI-007622	Applying	0.534	125.3	42.5	0.51	0.95 (0.96, 1.04)	-2.6	2420
OSAI-011447	Knowing	-0.422	115.8	67.8	0.31	1.12 (0.96, 1.04)	6.5	3769
OSAI-008375	Morphology	-0.761	112.4	73.3	0.44	0.98 (0.96, 1.04)	-1.2	3769
OSAI-008335	Applying	2.368	143.7	18.9	0.20	1.14 (0.95, 1.05)	5.0	3769
OSAI-008351	Categorising	0.384	123.8	52.8	0.29	1.17 (0.97, 1.03)	10.7	3769
OSAI-011429	Categorising	1.726	137.3	28.1	0.37	1.04 (0.96, 1.04)	1.9	3769
OSAI-011427	Categorising	1.982	139.8	24.2	0.45	0.91 (0.96, 1.04)	-4.1	3769
OSAI-011431	Categorising	1.484	134.8	32.2	0.46	0.95 (0.96, 1.04)	-2.7	3769
OSAI-011461	Categorising	1.529	135.3	31.4	0.35	1.06 (0.96, 1.04)	3.2	3769
OSAI-007600	Knowing	-1.236	107.6	78.1	0.50	0.86 (0.94, 1.06)	-4.6	2036
OSAI-007631	Applying	-1.490	105.1	76.6	0.54	1.02 (0.93, 1.07)	0.5	2036
OSAI-011432	Categorising	1.565	135.7	28.8	0.48	0.93 (0.95, 1.05)	-2.6	2036

Item label	Process	Item difficulty estimate (logit)	Item difficulty estimate (scale score)	Facility	Discrimination (Item–Rest corr.)	Weighted Fit		No. of data points
						Weighted mean square (95% confidence interval)	T value	
OSAI-008301	Applying	0.902	129.0	40.4	0.38	1.08 (0.96, 1.04)	3.5	2036
OSAI-011453	Morphology	0.533	125.3	47.4	0.33	1.15 (0.96, 1.04)	6.9	2036
OSAI-008369	Morphology	0.567	125.7	46.8	0.38	1.08 (0.96, 1.04)	3.9	2036
OSAI-008344	Categorising	1.161	131.6	35.7	0.47	0.98 (0.95, 1.05)	-1.0	2036
OSAI-008186	Applying	0.290	122.9	54.3	0.40	1.02 (0.96, 1.04)	1.0	1897
OSAI-008334	Applying	1.809	138.1	25.1	0.40	1.02 (0.94, 1.06)	0.6	2036
OSAI-008374	Applying	1.284	132.8	35.3	0.45	0.98 (0.95, 1.05)	-0.9	1888
OSAI-011434	Categorising	1.806	138.1	25.1	0.40	1.01 (0.94, 1.06)	0.2	2036
OSAI-011435	Categorising	4.089	160.9	4.6	0.27	0.95 (0.83, 1.17)	-0.5	2036
OSAI-011456	Morphology	-0.796	112.0	71.6	0.48	0.93 (0.95, 1.05)	-2.7	2036
OSAI-007619	Knowing	-0.122	118.8	59.9	0.50	0.93 (0.96, 1.04)	-3.3	2036
OSAI-007624	Knowing	1.230	132.3	33.3	0.46	0.95 (0.95, 1.05)	-2.3	2036
OSAI-007628	Knowing	1.880	138.8	24.6	0.44	0.94 (0.94, 1.06)	-1.9	2036
OSAI-007654	Knowing	0.860	128.6	45.9	0.36	1.14 (0.96, 1.04)	6.1	2036
OSAI-008299	Knowing	2.035	140.3	26.3	0.40	0.97 (0.96, 1.04)	-1.5	3145
OSAI-008303	Knowing	1.200	132.0	41.3	0.25	1.16 (0.97, 1.03)	9.4	3145
OSAI-008120	Categorising	0.030	120.3	64.2	0.39	1.00 (0.97, 1.03)	0.3	3145
OSAI-008167	Morphology	-0.236	117.6	69.0	0.48	0.91 (0.96, 1.04)	-4.5	3145
OSAI-008304	Categorising	1.322	133.2	39.0	0.35	1.05 (0.97, 1.03)	3.2	3145
OSAI-008199	Categorising	1.880	138.8	28.9	0.17	1.21 (0.96, 1.04)	9.6	3145
OSAI-008197	Categorising	0.813	128.1	49.0	0.49	0.92 (0.97, 1.03)	-5.7	3145
OSAI-008204	Categorising	2.039	140.4	26.3	0.35	1.01 (0.96, 1.04)	0.6	3145
OSAI-011455	Morphology	-0.624	113.8	75.3	0.42	0.93 (0.95, 1.05)	-2.9	3145
OSAI-008177	Categorising	1.239	132.4	42.6	0.44	0.98 (0.97, 1.03)	-1.4	4479
OSAI-007611	Knowing	-1.000	110.0	76.7	0.45	0.98 (0.93, 1.07)	-0.5	1733
OSAI-007663	Knowing	1.890	138.9	37.2	0.38	1.25 (0.94,1.06)	8.1	1733
OSAI-008187	Knowing	0.378	123.8	59.4	0.45	0.96 (0.97, 1.03)	-2.4	3067
OSAI-008134_78	Applying	2.588	145.9	18.3	0.14	1.18 (0.92, 1.08)	4.2	1653
OSAI-008134_910	Applying	2.328	143.3	29.4	0.18	1.18 (0.94, 1.06)	5.2	1204
OSAI-008181	Categorising	1.623	136.2	32.0	0.41	1.00 (0.95, 1.05)	0.0	1733
OSAI-008179_78	Categorising	2.168	141.7	23.3	0.52	0.86 (0.94, 1.06)	-4.4	1733
OSAI-008179_910	Categorising	1.999	140.0	32.8	0.57	0.84 (0.94, 1.06)	-6.0	1334
OSAI-008311	Categorising	0.780	127.8	48.0	0.38	1.05 (0.96, 1.04)	2.5	1733
OSAI-008188_78	Applying	0.114	121.1	61.0	0.38	1.04 (0.96, 1.04)	1.8	1733
OSAI-008188_910	Applying	0.655	126.6	58.8	0.41	1.00 (0.95, 1.05)	0.1	1334
OSAI-008111	Applying	1.189	131.9	40.0	0.32	1.12 (0.96, 1.04)	5.1	1733
OSAI-008108	Applying	1.165	131.6	40.5	0.21	1.23 (0.96, 1.04)	9.5	1733
OSAI-008122	Applying	0.168	121.7	60.0	0.48	0.93 (0.96, 1.04)	-3.0	1733
OSAI-008308	Applying	2.308	143.1	24.0	0.23	1.13 (0.95, 1.05)	5.4	3067
OSAI-007680	Applying	0.237	122.4	55.2	0.51	0.94 (0.96, 1.04)	-3.0	1733
OSAI-007676	Knowing	0.719	127.2	49.2	0.40	1.03 (0.96, 1.04)	1.4	1733
OSAI-007671	Knowing	-0.550	114.5	70.2	0.57	0.86 (0.94, 1.06)	-5.1	1733
OSAI-007652	Applying	0.670	126.7	47.6	0.52	0.91 (0.96, 1.04)	-4.4	1733
OSAI-007672	Knowing	-0.709	112.9	75.3	0.49	0.87 (0.94, 1.06)	-4.3	1733
OSAI-011430	Categorising	1.557	135.6	38.6	0.42	0.96 (0.97, 1.03)	-2.3	2746
OSAI-011463	Categorising	1.061	130.6	48.3	0.24	1.18 (0.97, 1.03)	10.8	2746
OSAI-011410	Categorising	1.651	136.5	36.8	0.34	1.05 (0.96, 1.04)	2.8	2746

Item label	Process	Item difficulty estimate (logit)	Item difficulty estimate (scale score)	Facility	Discrimination (Item–Rest corr.)	Weighted Fit		No. of data points
						Weighted mean square (95% confidence interval)	T value	
OSAI-011411	Categorising	1.692	136.9	36.0	0.29	1.10 (0.96, 1.04)	5.1	2746
OSAI-011418	Applying	1.978	139.8	30.8	0.27	1.10 (0.96, 1.04)	4.7	2746
OSAI-011422	Applying	2.194	141.9	27.2	0.30	1.07 (0.96, 1.04)	3.0	2746
OSAI-011445_89	Categorising	2.625	146.2	18.8	0.27	1.05 (0.92, 1.08)	1.2	1412
OSAI-011445_910	Categorising	3.280	152.8	14.5	0.28	1.04 (0.90, 1.10)	0.7	1334
OSAI-011450	Categorising	0.077	120.8	67.3	0.50	0.88 (0.96, 1.04)	-6.4	2746
OSAI-011452	Morphology	-1.339	106.6	86.8	0.44	0.82 (0.92, 1.08)	-4.7	2746
OSAI-007632	Knowing	-0.158	118.4	69.6	0.45	0.92 (0.94, 1.06)	-2.7	1412
OSAI-008340	Morphology	1.043	130.4	46.3	0.36	1.04 (0.96, 1.04)	1.6	1412
OSAI-008332	Applying	0.073	120.7	68.2	0.39	0.96 (0.95, 1.05)	-1.5	1327
OSAI-008377	Categorising	2.993	149.9	14.5	0.23	1.07 (0.90, 1.10)	1.3	1412
OSAI-008283	Morphology	-0.050	119.5	67.7	0.45	0.92 (0.94, 1.06)	-2.7	1412
OSAI-008338	Applying	1.050	130.5	46.2	0.48	0.93 (0.96, 1.04)	-3.2	1412
OSAI-008373	Applying	2.469	144.7	21.0	0.18	1.12 (0.92, 1.08)	3.0	1412
OSAI-008371	Applying	0.337	123.4	60.5	0.37	1.03 (0.95, 1.05)	1.0	1412
OSAI-008336	Knowing	2.370	143.7	22.4	0.18	1.14 (0.93, 1.07)	3.6	1412
OSAI-008286	Knowing	1.655	136.6	34.4	0.35	1.03 (0.95, 1.05)	1.2	1412
OSAI-011428	Categorising	2.390	143.9	22.1	0.43	0.92 (0.93, 1.07)	-2.2	1412
OSAI-011451	Knowing	-0.297	117.0	72.0	0.45	0.92 (0.94, 1.06)	-2.6	1412
OSAI-007664	Applying	2.000	140.0	26.1	0.28	1.03 (0.94, 1.06)	1.1	1412
OSAI-007684	Applying	0.486	124.9	45.3	0.40	1.11 (0.95, 1.05)	4.4	1412
OSAI-007705	Applying	1.973	139.7	28.7	0.35	1.02 (0.94, 1.06)	0.6	1412
OSAI-007706	Applying	1.903	139.0	29.9	0.32	1.05 (0.94, 1.06)	1.6	1412
OSAI-008178	Categorising	2.381	143.8	26.4	0.53	0.85 (0.93, 1.07)	-4.7	1334
OSAI-008319	Applying	3.564	155.6	12.2	0.34	0.96 (0.88, 1.12)	-0.7	1268
OSAI-011424	Applying	2.723	147.2	21.3	0.23	1.11 (0.92, 1.08)	2.8	1334
OSAI-007697	Applying	1.167	131.7	48.7	0.40	1.02 (0.95, 1.05)	0.9	1334
OSAI-007734	Applying	0.977	129.8	55.2	0.42	1.00 (0.95, 1.05)	0.0	1334
OSAI-008328	Categorising	1.463	134.6	42.9	0.35	1.08 (0.95, 1.05)	3.1	1334
OSAI-008142	Categorising	3.323	153.2	14.0	0.32	0.99 (0.89, 1.11)	-0.2	1334
OSAI-008172	Applying	2.141	141.4	30.4	0.37	1.02 (0.94, 1.06)	0.8	1334
OSAI-008309	Categorising	1.579	135.8	40.6	0.39	1.03 (0.95, 1.05)	1.1	1334
OSAI-008312	Categorising	2.638	146.4	22.5	0.18	1.18 (0.93, 1.07)	4.4	1334
OSAI-008176	Categorising	0.306	123.1	65.3	0.58	0.80 (0.94, 1.06)	-7.4	1334
OSAI-008162	Categorising	3.650	156.5	10.9	0.36	0.93 (0.88, 1.12)	-1.1	1334
OSAI-008140	Morphology	2.547	145.5	23.8	0.40	0.98 (0.93, 1.07)	-0.4	1334
OSAI-008168	Morphology	2.992	149.9	17.8	0.33	1.01 (0.91, 1.09)	0.3	1334
OSAI-008115	Categorising	0.209	122.1	67.0	0.46	0.91 (0.94, 1.06)	-3.1	1334
OSAI-008141	Categorising	1.630	136.3	39.7	0.32	1.11 (0.95, 1.05)	4.1	1334
OSAI-008171	Knowing	0.574	125.7	64.1	0.28	1.08 (0.95, 1.05)	2.8	1205
OSAI-007765	Applying	1.630	136.3	44.9	0.42	1.06 (0.95, 1.05)	2.3	1334
OSAI-007687	Applying	0.860	128.6	58.1	0.43	0.98 (0.95, 1.05)	-1.0	1334
OSAI-007752	Applying	1.250	132.5	49.6	0.53	0.90 (0.95, 1.05)	-4.4	1334
OSAI-007814	Applying	2.170	141.7	33.0	0.55	0.91 (0.94, 1.06)	-3.1	1334
OSAI-007854	Applying	2.765	147.7	21.3	0.27	1.11 (0.92, 1.08)	2.7	1334

Appendix 2

Scale score tranformations

The *PAT Vocabulary Skills* tests were calibrated using the Rasch model. The estimates of student ability in logits were computed based on observed raw scores in each test, and equated onto the corresponding logit scale. For reporting purposes, the estimates of ability for each test were transformed onto scale scores. The unit used to express scale scores is defined from the Rasch measurement unit, the logit: 1 logit = 10 *PAT Vocabulary Skills* scale scores. This has been done to avoid assigning negative values to performance measures. The formula for transforming *PAT Vocabulary Skills* logit values to scale score values are as follows:

Scale score of item estimate = logit *10 + 120.

Scale score of ability estimate = logit *10 + 120.

Appendix 3

Score equivalence tables for PAT Vocabulary Skills final test forms

PAT Vocabulary Skills Test 1

Raw score	Ability estimate (logit)	Standard error (logit)	Ability estimate (scale score)	SE (scale score)	Band
0	-7.094	1.480	49.1	14.8	Band 1
1	-5.930	0.880	60.7	8.8	Band 1
2	-5.352	0.701	66.5	7.0	Band 1
3	-4.946	0.610	70.5	6.1	Band 1
4	-4.623	0.554	73.8	5.5	Band 1
5	-4.349	0.516	76.5	5.2	Band 1
6	-4.105	0.489	78.9	4.9	Band 1
7	-3.884	0.469	81.2	4.7	Band 1
8	-3.677	0.454	83.2	4.5	Band 1
9	-3.481	0.443	85.2	4.4	Band 1
10	-3.294	0.435	87.1	4.4	Band 1
11	-3.112	0.429	88.9	4.3	Band 1
12	-2.934	0.426	90.7	4.3	Band 2
13	-2.758	0.424	92.4	4.2	Band 2
14	-2.582	0.424	94.2	4.2	Band 2
15	-2.406	0.426	95.9	4.3	Band 2
16	-2.228	0.429	97.7	4.3	Band 2
17	-2.046	0.435	99.5	4.4	Band 2
18	-1.859	0.443	101.4	4.4	Band 3
19	-1.663	0.454	103.4	4.5	Band 3
20	-1.457	0.469	105.4	4.7	Band 3
21	-1.236	0.489	107.6	4.9	Band 3
22	-0.993	0.516	110.1	5.2	Band 4
23	-0.720	0.553	112.8	5.5	Band 4
24	-0.398	0.609	116.0	6.1	Band 4
25	0.006	0.700	120.1	7.0	Band 5
26	0.582	0.878	125.8	8.8	Band 5
27	1.744	1.479	137.4	14.8	Band 6
28	1.336	0.693	133.4	6.9	Band 6
29	1.906	0.872	139.1	8.7	Band 6
30	3.063	1.473	150.6	14.7	Band 8
23	-0.072	0.452	119.3	4.5	Band 4
24	0.135	0.474	121.4	4.7	Band 5
25	0.365	0.503	123.7	5	Band 5
26	0.628	0.542	126.3	5.4	Band 5
27	0.940	0.600	129.4	6	Band 5

PAT Vocabulary Skills Test 2

Raw score	Ability estimate (logit)	Standard error (logit)	Ability estimate (scale score)	SE (scale score)	Band
0	-5.809	1.469	61.9	14.7	Band 1
1	-4.656	0.869	73.4	8.7	Band 1
2	-4.091	0.689	79.1	6.9	Band 1
3	-3.698	0.597	83.0	6	Band 1
4	-3.390	0.539	86.1	5.4	Band 1
5	-3.130	0.500	88.7	5	Band 1
6	-2.903	0.471	91.0	4.7	Band 2
7	-2.698	0.450	93.0	4.5	Band 2
8	-2.509	0.434	94.9	4.3	Band 2
9	-2.332	0.421	96.7	4.2	Band 2
10	-2.164	0.411	98.4	4.1	Band 2
11	-2.002	0.404	100.0	4	Band 3
12	-1.845	0.399	101.5	4	Band 3
13	-1.692	0.395	103.1	4	Band 3
14	-1.540	0.393	104.6	3.9	Band 3
15	-1.390	0.393	106.1	3.9	Band 3
16	-1.239	0.393	107.6	3.9	Band 3
17	-1.088	0.396	109.1	4	Band 3
18	-0.933	0.400	110.7	4	Band 4
19	-0.776	0.405	112.2	4.1	Band 4
20	-0.613	0.413	113.9	4.1	Band 4
21	-0.443	0.423	115.6	4.2	Band 4
22	-0.264	0.436	117.4	4.4	Band 4
23	-0.072	0.452	119.3	4.5	Band 4
24	0.135	0.474	121.4	4.7	Band 5
25	0.365	0.503	123.7	5	Band 5
26	0.628	0.542	126.3	5.4	Band 5
27	0.940	0.600	129.4	6	Band 5
28	1.336	0.693	133.4	6.9	Band 6
29	1.906	0.872	139.1	8.7	Band 6
30	3.063	1.473	150.6	14.7	Band 8

PAT Vocabulary Skills Test 3

Raw score	Ability estimate (logit)	Standard error (logit)	Ability estimate (scale score)	SE (scale score)	Band
0	-4.660	1.462	73.4	14.6	Band 1
1	-3.514	0.863	84.9	8.6	Band 1
2	-2.955	0.683	90.4	6.8	Band 2
3	-2.570	0.590	94.3	5.9	Band 2
4	-2.269	0.531	97.3	5.3	Band 2
5	-2.017	0.491	99.8	4.9	Band 2
6	-1.798	0.462	102.0	4.6	Band 3
7	-1.601	0.440	104.0	4.4	Band 3
8	-1.420	0.423	105.8	4.2	Band 3
9	-1.253	0.409	107.5	4.1	Band 3
10	-1.094	0.398	109.1	4	Band 3
11	-0.943	0.390	110.6	3.9	Band 4
12	-0.797	0.383	112.0	3.8	Band 4
13	-0.656	0.378	113.4	3.8	Band 4
14	-0.517	0.375	114.8	3.7	Band 4
15	-0.381	0.373	116.2	3.7	Band 4
16	-0.245	0.372	117.5	3.7	Band 4
17	-0.110	0.372	118.9	3.7	Band 4
18	0.025	0.373	120.3	3.7	Band 5
19	0.162	0.376	121.6	3.8	Band 5
20	0.301	0.379	123.0	3.8	Band 5
21	0.443	0.385	124.4	3.8	Band 5
22	0.590	0.391	125.9	3.9	Band 5
23	0.743	0.400	127.4	4	Band 5
24	0.903	0.411	129.0	4.1	Band 5
25	1.073	0.425	130.7	4.2	Band 6
26	1.255	0.442	132.6	4.4	Band 6
27	1.454	0.464	134.5	4.6	Band 6
28	1.676	0.494	136.8	4.9	Band 6
29	1.931	0.534	139.3	5.3	Band 6
30	2.235	0.592	142.3	5.9	Band 7
31	2.623	0.685	146.2	6.9	Band 7
32	3.185	0.865	151.9	8.7	Band 8
33	4.334	1.465	163.3	14.7	Band 8

PAT Vocabulary Skills Test 4

Raw score	Ability estimate (logit)	Standard error (logit)	Ability estimate (scale score)	SE (scale score)	Band
0	-3.503	1.458	85.0	14.6	Band 1
1	-2.362	0.859	96.4	8.6	Band 2
2	-1.808	0.679	101.9	6.8	Band 3
3	-1.427	0.585	105.7	5.9	Band 3
4	-1.130	0.527	108.7	5.3	Band 3
5	-0.882	0.487	111.2	4.9	Band 4
6	-0.667	0.457	113.3	4.6	Band 4
7	-0.474	0.435	115.3	4.3	Band 4
8	-0.297	0.418	117.0	4.2	Band 4
9	-0.134	0.404	118.7	4	Band 4
10	0.021	0.393	120.2	3.9	Band 5
11	0.168	0.384	121.7	3.8	Band 5
12	0.309	0.377	123.1	3.8	Band 5
13	0.447	0.372	124.5	3.7	Band 5
14	0.580	0.368	125.8	3.7	Band 5
15	0.712	0.365	127.1	3.7	Band 5
16	0.842	0.363	128.4	3.6	Band 5
17	0.971	0.363	129.7	3.6	Band 5
18	1.099	0.363	131.0	3.6	Band 6
19	1.229	0.364	132.3	3.6	Band 6
20	1.359	0.366	133.6	3.7	Band 6
21	1.492	0.370	134.9	3.7	Band 6
22	1.627	0.374	136.3	3.7	Band 6
23	1.766	0.380	137.7	3.8	Band 6
24	1.910	0.387	139.1	3.9	Band 6
25	2.060	0.396	140.6	4	Band 7
26	2.217	0.408	142.2	4.1	Band 7
27	2.385	0.422	143.8	4.2	Band 7
28	2.565	0.439	145.6	4.4	Band 7
29	2.762	0.462	147.6	4.6	Band 7
30	2.981	0.491	149.8	4.9	Band 7
31	3.234	0.532	152.3	5.3	Band 8
32	3.536	0.590	155.4	5.9	Band 8
33	3.922	0.683	159.2	6.8	Band 8
34	4.482	0.864	164.8	8.6	Band 8
35	5.629	1.463	176.3	14.6	Band 8

PAT Vocabulary Skills Test 5

Raw score	Ability estimate (logit)	Standard error (logit)	Ability estimate (scale score)	SE (scale score)	Band
0	-2.550	1.453	94.5	14.5	Band 2
1	-1.414	0.854	105.9	8.5	Band 3
2	-0.864	0.674	111.4	6.7	Band 4
3	-0.488	0.580	115.1	5.8	Band 4
4	-0.196	0.521	118.0	5.2	Band 4
5	0.046	0.480	120.5	4.8	Band 5
6	0.255	0.450	122.5	4.5	Band 5
7	0.441	0.427	124.4	4.3	Band 5
8	0.611	0.409	126.1	4.1	Band 5
9	0.768	0.394	127.7	3.9	Band 5
10	0.915	0.383	129.2	3.8	Band 5
11	1.055	0.373	130.5	3.7	Band 6
12	1.188	0.366	131.9	3.7	Band 6
13	1.316	0.359	133.2	3.6	Band 6
14	1.441	0.354	134.4	3.5	Band 6
15	1.562	0.351	135.6	3.5	Band 6
16	1.681	0.348	136.8	3.5	Band 6
17	1.799	0.346	138.0	3.5	Band 6
18	1.916	0.345	139.2	3.5	Band 6
19	2.033	0.345	140.3	3.4	Band 7
20	2.149	0.346	141.5	3.5	Band 7
21	2.267	0.347	142.7	3.5	Band 7
22	2.386	0.349	143.9	3.5	Band 7
23	2.506	0.353	145.1	3.5	Band 7
24	2.630	0.357	146.3	3.6	Band 7
25	2.756	0.362	147.6	3.6	Band 7
26	2.887	0.369	148.9	3.7	Band 7
27	3.023	0.377	150.2	3.8	Band 8
28	3.166	0.387	151.7	3.9	Band 8
29	3.317	0.399	153.2	4	Band 8
30	3.478	0.414	154.8	4.1	Band 8
31	3.653	0.432	156.5	4.3	Band 8
32	3.844	0.455	158.4	4.6	Band 8
33	4.059	0.486	160.6	4.9	Band 8
34	4.308	0.527	163.1	5.3	Band 8
35	4.606	0.586	166.1	5.9	Band 8
36	4.989	0.680	169.9	6.8	Band 8
37	5.547	0.861	175.5	8.6	Band 8
38	6.691	1.462	186.9	14.6	Band 8