



NEW ZEALAND **QUALIFICATIONS** AUTHORITY
MANA TOHU MĀTAURANGA O AOTEAROA

QUALIFY FOR THE FUTURE WORLD
KIA NOHO TAKATŪ KI TŌ ĀMUA AO!

Psychometric and Statistical Analysis – English Level 1 Externally Assessed Achievement Standards Using Digital Medium

Co-authored by Dr Michael Johnston, Senior Lecturer, School
of Education, Victoria University & Eldon Paki, Senior
Statistical Analyst, NZQA

Psychometric and statistical analysis - English Level 1 externally-assessed achievement standards using digital medium

Introduction

This report presents statistical analyses comparing the psychometric properties of external assessments for Level 1 English presented and completed in digital format, with those of the same assessments presented and completed in paper format. The data supporting these analyses were the externally-assessed Level 1 English results for all students from the 35 schools that participated in the 2016 Pilot of the digital assessment format. While there were sufficient results from the digital format to support a reliable analysis, there were approximately four times as many results at these schools from assessments completed on paper format as there were from those completed digitally. These paper-format results from the participating schools provided the data against which to compare the digital-format data in the present analyses.

Summary of Findings

There were some statistically significant differences between the results distributions for digital- and paper-format assessments. Specifically, for two of the three Level 1 English standards in the digital Pilot, percentages of Achieved results were significantly higher for paper-format assessments than for digital-format assessments, and percentages of Excellence results were significantly and commensurately lower. Rasch analyses showed significant differences in difficulty parameters that were consistent with the differences in the percentage-distributions of results.

In interpreting these differences, it is important to consider that the groups of students undertaking assessments in each of the digital and paper formats were self-selecting and that there is therefore no basis to assume that the two groups were equal in ability, or that they ought to have attained the same distributions of results. Indeed there is some evidence, albeit relatively weak, to suggest that more confident – and therefore possibly more able – students might have been more likely than less confident (and able) students to opt for digital-format assessment (see commentary on Table 1 below). In any event, the differences in percentages of students attaining Achieved and Excellence and the associated differences in the Rasch model parameters cannot be taken as reliable evidence that either format provided a performance advantage.

A pair of regression analyses investigating the relationship between internal assessment and external assessments under each of the digital and paper formats showed no differences in model parameters. Thus, the present results show no evidence that the validity of the two assessments formats differ, if the validity of external assessment is operationalised in terms of its predictive relationship with internal assessment.

Results

Most of the 35 participating schools had results for all three externally-assessed Level 1 English standards: 90849, 90850, and 90851 and most had at least some students completing each standard in each of the digital and paper formats.

Comparison of grade distributions for digital and paper formats

Table 1 shows the number and name of each externally-assessed achievement standard for Level 1 English, as well as the total numbers of results for each of the paper and digital formats at the participating schools, and the proportions of all results that were for students completing each standard in digital format. While similar percentages (between 16.7% and 20.6%) of candidates for each standard used digital format, the standard with the fewest results overall (90851) had the least proportion of candidates undertaking it in digital format, and the standard with the greatest number of results overall (90850) had the greatest proportion of candidates undertaking it in digital format. A plausible explanation for this is that students might have been less likely to opt to use the digital format for standards for which they felt less confident. (This interpretation rests on an assumption that greater numbers of results reflect greater student confidence.)

Table 1: Total numbers of results for Level 1 externally-assessed achievement standards in English at the 34 schools participating in the digital assessment Pilot

Standard Number	Standard Title	Total results: Digital format	Total results: Paper format	Digital format results as percentages of all results
90849	Show understanding of specified aspect(s) of studied written text(s), using supporting evidence	742	3,183	18.9
90850	Show understanding of specified aspect(s) of studied visual or oral text(s), using supporting evidence	835	3,221	20.6
90851	Show understanding of significant aspects of unfamiliar written text(s) through close reading, using supporting evidence	587	2,919	16.7

Figure 1 compares the distributions of grades for digital and paper formats for each of the three Level 1 externally-assessed achievement standards in English. For all three standards, the percentage of *Not Achieved* and *Achieved* grades was higher for paper format than for digital format, and the percentages of *Merit* and *Excellence* grades were commensurately lower. Z tests showed that the differences between the two formats in the percentages of *Achieved* and *Excellence* grades were statistically significant ($p < .05$) for standards 90849 and 90850; no other differences were significant.

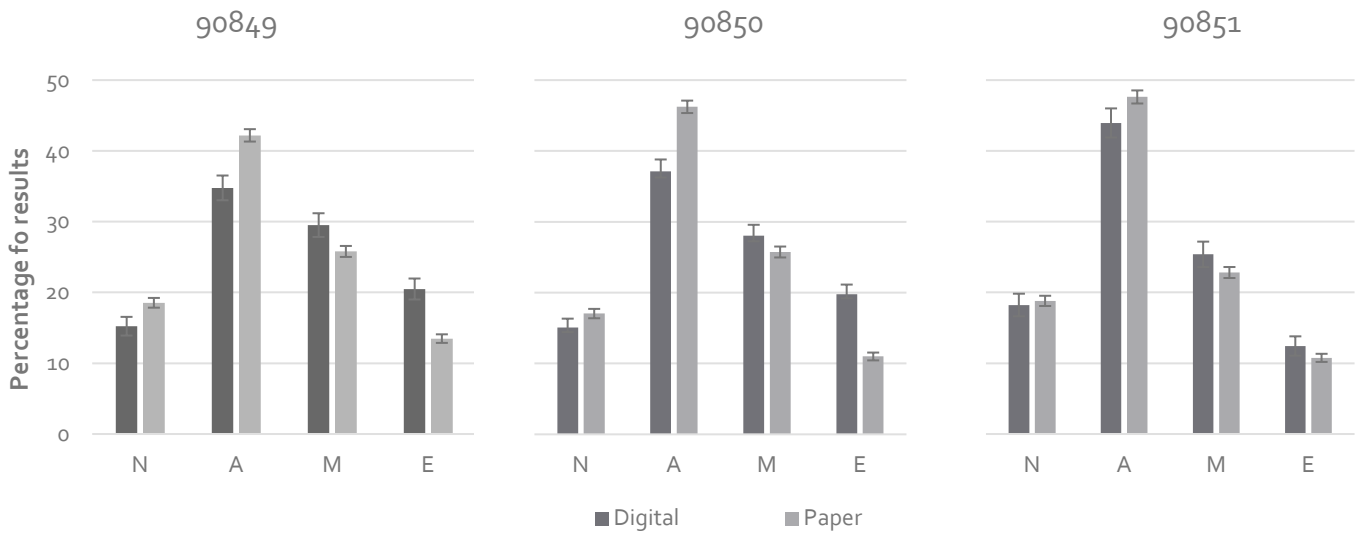


Figure 1.: Comparisons of digital and paper results distributions for Level 1 externally-assessed achievement standards in English at schools participating in the digital Pilots. Vertical bars denote standard errors.

It is important to note that the observed differences in the results distributions for the two formats might be attributable to differences in relative capability between candidates completing assessments in the digital format and those completing them in paper format. In particular, more capable students might, on average, have felt more confident to use the digital format than less capable students. This would be consistent with the data in Table 1 showing that the proportion of students undertaking digital assessment in each standard was correlated with the total number of results: Students may have been more likely to undertake digital assessment for standards for which they felt relatively confident.

Full results distributions for each standard and each school, for each of the digital and paper formats are shown in an appendix.

Comparison of Rasch difficulty parameters for digital and paper formats

Subsets of students were identified who completed assessments only in digital format and only in paper format. Three Rasch analyses were conducted for each of these subsets to identify difficulty parameters associated with grades of *Achieved* or better, grades of *Merit* or *Excellence* and grades of *Excellence*. In each of these analyses, the individual standards were treated as items, so that the measurement scale on which the difficulty parameters were estimated reflected aggregated performance across the three standards, allowing for a quantitative (interval scale) comparison of difficulty. Figure 2 shows comparisons of these parameters for each of the three standards.

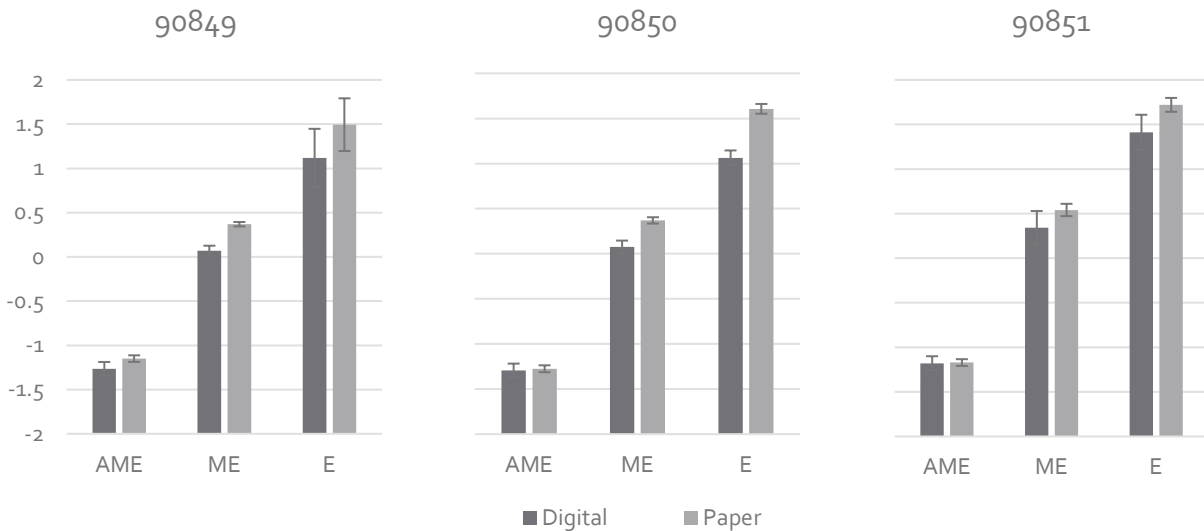


Figure 2.: Comparisons of Rasch difficulty parameters for attaining grades of *Achieved or better* (AME), *Merit or better* (ME) and *Excellence* (E), for digital and paper formats for each externally-assessed Level 1 achievement standard in English. Vertical bars denote standard errors.

For all three standards the difficulty parameters associated with attaining grades of *Achieved or better* were very similar for digital and paper formats, but the parameters associated with grades of *Merit or better* and with grades of *Excellence* were somewhat higher for the paper format than for the digital format. The differences between the pairs of parameter values for the two formats exceeded the 95% confidence interval for the difference in the case of ME for standard 90849, and for ME and E for standard 90850. These differences may therefore be treated as being statistically significant. No other differences between pairs of parameters exceeded the 95% confidence interval for the difference.

Analysis of the relative validity of digital and paper formats

This section of the report investigates the extent to which the results of the digital and paper mediums are equivalent in terms of their predictive relationships with internally-assessed achievement standards at Level 1. This constitutes a limited check on the relative validity of the two assessment formats; any significant difference in the predictive relationship with internal assessment could comprise an associated difference in validity.

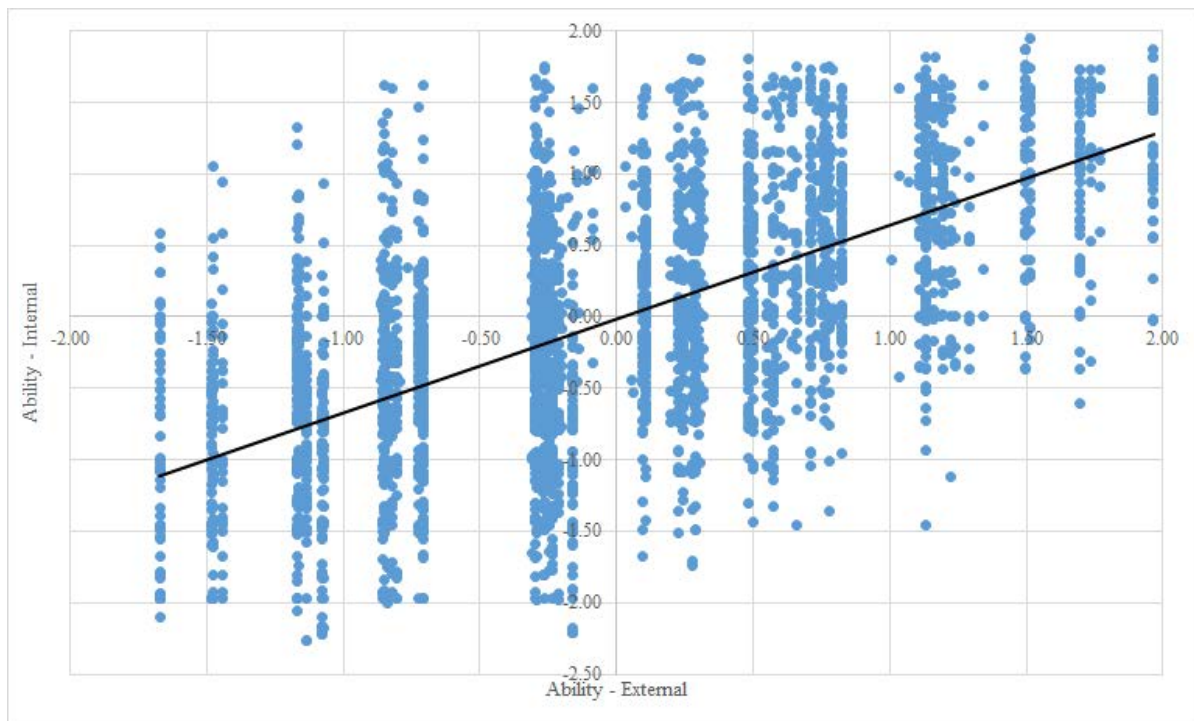
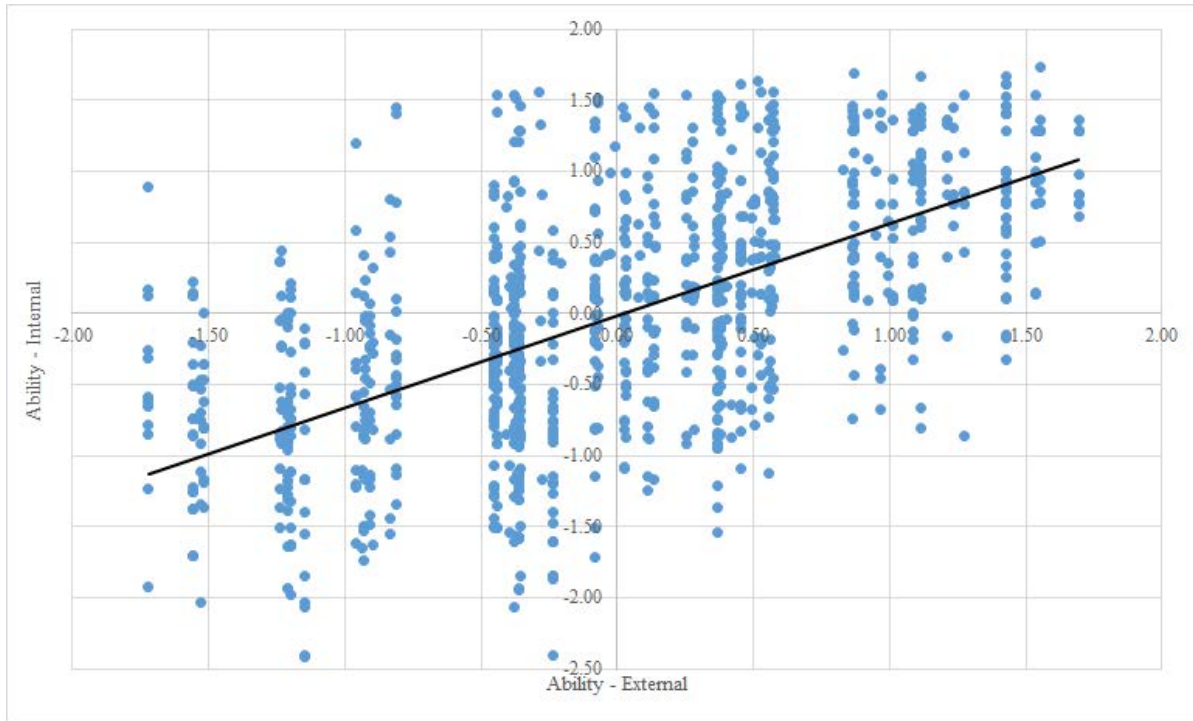
The performance variables for the two external assessment formats comprised the ability estimates for each candidate resulting from the Rasch analysis described above. A third Rasch analysis was carried out on the internally-assessed results of both the students who completed external assessments digitally and those who completed them on paper. Like the analyses of the two external assessment formats, this analysis treated individual internally-assessed standards as items, yielding an interval-scale measurement variable as an aggregate measure of performance across internally-assessed standards.

A least-squares linear regression analysis was used to model the relationship between digital-format external assessment and internal assessment and another to model the relationship between paper-format external assessment and internal assessment. Figure 3 depicts two

scatterplots, with regression lines, showing the relationship between digital-format external assessment and internal assessment (upper panel) and paper-format external assessment and internal assessment (lower panel).

A comparison of the constant and slope parameters estimated under regression models for each external assessment format allows for a statistical comparison of the equivalence of the predictive relationships. These comparisons showed that the differences between both the constants and the slope parameters were well inside the 95% confidence intervals for their respective differences. Thus there is no evidence that the two formats differed in respect of their validity, in terms of the predictive relationship between external and internal assessment.

Figure 3.: Scatterplots with regression lines showing the relationships between Rasch ability parameters estimated for external assessment and internal assessment. The upper panel shows the relationship for digital-format external assessment and the lower panel the relationship for paper-format external assessment.



APPENDIX

Numbers of results in each grade category for standard 90849 for each participating school by medium of delivery.

		90849							
		Digital				Paper			
School	Decile	N	A	M	E	N	A	M	E
1	9		5	11	4	40	144	45	11
2	4	5	12	14	6	60	74	18	2
3	8			1	2	57	104	64	26
4	4	2	11	1	5	52	71	24	19
5	8	4	7	3	3	22	77	56	23
6	9	3	22	36	19	11	87	88	50
7	10	1	18	8	17	4	21	41	26
8	10	1	7	5	8	3	27	51	57
9	10	3	33	64	46	4	15	30	8
10	1	2							
11	5	1	6	5	1	78	83	35	18
12	6	1	12	9	1	3	12	8	3
13	3	2	5						
14	7					14	54	49	21
15	9	1	1	2		29	84	46	9
16	3	18	1	1		8	3		
17	1	6	3	1		2	1	2	
18	6	1	3	1		56	77	22	12
19	5		3	1	1	7	26	16	4
20	4					9	30	11	6
21	4	15	6	2	2	12	15	6	1
22	8	1	13	7	3	13	36	25	16
23	1	15	11	1		2	2		
24	6	1				1	1	1	2
25	4	11	16	4	2	4	1	1	
26	6					12	86	35	6
27	9	1	1	5	5	8	30	23	29

28	7		1		1	17	31	22	23
29	10	15	29	17	10	10	15	18	11
30	8	1	4	5	1	28	72	46	25
31	10		4	3		7	25	17	12
32	6	1	13	8	4		6	3	2
33	9		3	3	10			2	1
34	2								
35	10	1	8	1	1	17	33	16	6
Total		11 3	25 8	21 9	15 2	59 0	1,343	82 1	42 9
Proportion by medium of delivery		.15	.35	.30	.21	.19	.42	.26	.14

Numbers of results in each grade category for standard 90850 for each participating school by medium of delivery.

		90850							
		Digital				Paper			
School	Decile	N	A	M	E	N	A	M	E
1	9		8	6	4	29	113	54	10
2	4	6	12	12	11	67	98	34	5
3	8			2	1	38	112	64	18
4	4	1	12	5	2	26	57	36	14
5	8	1	5	6	5	22	88	51	22
6	9	2	32	20	21	5	79	52	49
7	10	1	15	12	16	3	29	44	18
8	10	3	3	8	7	9	37	49	48
9	10	4	43	48	45	1	23	29	8
10	1					1	2	1	
11	5	2	6	7	1	67	117	35	13
12	6	3	11	7	2	18	27	17	3
13	3	6	17	8	1	1	1		
14	7					17	65	52	20
15	9	1	1	2		30	96	47	13
16	3	18	4	1		7	5		
17	1	3							
18	6	1	2	2		45	102	39	14
19	5		3	1	1	10	27	16	4
20	4	8	16	2		15	32	20	5
21	4	16	8	2		13	13	4	5
22	8	8	21	13	9	19	60	35	22
23	1	15	8	2		7	1		
24	6	1		1	1	6	15	5	3
25	4	7	14	7	2	4	3		
26	6		6	5		14	77	25	8
27	9			1		10	27	8	5
28	7								

29	10	11	26	22	18	13	21	12	8
30	8	1	3	5	2	21	63	58	23
31	10	1	4	4	3	9	25	15	3
32	6	1	10	13	6	13	36	6	1
33	9	1	10	8	6		1	1	1
34	2	2	3						
35	10	2	7	2	1	9	37	20	11
Total		12 6	31 0	23 4	16 5	54 9	1,489	82 9	35 4
Proportion by medium of delivery		.15	.37	.28	.20	.17	.46	.26	.11

Numbers of results in each grade category for standard 90851 for each participating school by medium of delivery.

		90851							
		Digital				Paper			
School	Decile	N	A	M	E	N	A	M	E
1	9		7	9	1	44	120	43	6
2	4	1	7	14	9	81	95	17	5
3	8								
4	4	3	8	6	4	30	65	23	8
5	8	1	7	3	1	22	83	47	6
6	9	3	28	18	15	3	66	50	60
7	10	1	8	12	14	5	25	35	40
8	10		5	14	2	6	36	60	41
9	10	4	9	5	1		4	6	
10	1	5	2			36	34	1	
11	5	2	9	4		72	147	48	14
12	6		1						
13	3	13	18	6	3	2	1		
14	7			4	2	17	83	40	15
15	9					15	36	12	5
16	3	10	7	1		4	4		1
17	1								
18	6	1	5			56	95	33	12
19	5	1	3		1	7	28	17	5
20	4					17	57	24	7
21	4	10	9	3	3	13	20	6	3
22	8	10	19	9	4	15	54	33	15
23	1	2	2	1		1			
24	6		1	1		4	22	12	
25	4	25	45	10	3	3	14		2
26	6					1	13	14	1
27	9		3	6	3	15	63	46	21
28	7		1	1		23	47	17	7

29	10	12	39	16	5	8	26	12	7
30	8	1	6	3	1	24	88	39	18
31	10		2	1		13	21	20	6
32	6								
33	9								
34	2								
35	10	2	7	2	1	12	43	11	9
Total		107	25 8	14 9	73	54 9	1,390	666	31 4
Proportion by medium of delivery		.18	.44	.25	.12	.19	.48	.23	.11