Psychometric and statistical analysis of the 2018 Pilot delivery of English Level 1 and 2 externally assessed achievement standards using digital medium

Executive summary

Statistical analyses comparing the psychometric properties of externally-assessed standards completed in digital and paper formats were performed to investigate the extent to which the two formats afforded students equivalent assessment opportunities. Such equivalence is important to establish as NZQA expands the range of NCEA examinations available digitally.

The analyses were restricted to Level 1 and 2 English, as these two examinations had a large enough number of participants to provide reliable comparisons of the two formats. Digital- and paper-format results from the same set of schools participating in Digital Pilots were compared. Subsequent analysis compared the results after adjusting for student ability, by matching on students' internally-assessed standards. Both overall grade distributions and Rasch difficulty parameters of the two sets of results were examined.

Overall there was no conclusive evidence of a difference between the two formats of the assessment.

Introduction

NZQA has a goal of having NCEA examinations available online and, since 2016, has been working with schools to offer Digital Pilot examinations in preparation for this. 2018 Digital Pilot examinations contained the same content as the paper examinations (substitution), but some digital tools featured in the examinations such as spell check. Involvement in a Digital Pilot examination was voluntary, and those students who opted in sat a digital examination rather than the paper examination (the printed examination paper was available as a back up). The Digital Pilot examinations were held on the same dates and times as the paper-based examinations [1].

The purpose of the analyses in this report is to investigate the extent to which the two formats afforded students equivalent assessment opportunities. Such equivalence is important to establish as NZQA expands the digital assessment opportunities it offers to students.

A total of nine Digital Pilot examinations were offered in 2018 – Levels 1, 2, and 3 for English, Media Studies, and Classical Studies. This report analyses the results of Levels 1 and 2 English, as these two examinations had a large enough number of participants to provide reliable estimates of the psychometric properties of their respective external standards.

Students may be awarded a derived grade[2] if they believe their preparation for, or performance in, their examination was impaired. In order to reliably compare the two modes of external assessment, no student with a derived final grade was included in this analysis.

This report uses methodology from last year's analysis of 2017 results[3], and the same set of analyses was performed to evaluate Level 1 and Level 2 English examination results. The preliminary analysis compared students who sat the examination digitally with students who sat the same examination on paper, from the same schools. For each standard, the overall grade distributions (Not Achieved,

[1] For more information, refer to https://www.nzqa.govt.nz/about-us/future-state/digital-assessment-trials-pilots/about-2018/

[2] For more information, refer to https://www.nzqa.govt.nz/providers-partners/assessment-and-moderation-of-standards/managing-national-assessment-in-schools/derived-grades/

[3] For more information, refer to https://www.nzqa.govt.nz/assets/About-us/Future-State/2017-trials-and-pilots/Psychometricanalysesreport.pdf

Achieved, Merit, and Excellence) of these two sets of students were compared. Rasch analyses were also conducted to estimate the difficulty parameters associated with grades of Achieved or better, Merit or better, and Excellence for the two sets of students. This analysis treated individual externally assessed standards as items, yielding an interval-scale measurement variable as an aggregate measure of performance across externally assessed standards.

Students participating in the digital assessment were self-selecting. To control for the probable difference in ability between the two groups, another pair of datasets consisting of matched digital-format and paper-format students was constructed by matching on internally assessed achievement standards at the same level.

For each digital-format student, a matching paper-format student was randomly selected from the set of all paper-format students with the same profile of internal assessment results – that is, from those paper-format students who undertook the same set of internally-assessed standards and attained the same result for each – as the target digital-format student. A large number of resamples from the matching paper-format students was taken. As before, the overall grade distributions and difficulty parameters for Rasch analysis for the matched digital-format students and aggregate of matched paper-format students were compared. Any residual differences could then be attributed to the characteristic of the two examination formats.

1. Level 1 English

1.1 Students in schools participating in Digital Pilot

Table 1 shows the externally-assessed standards in Level 1 English as well as the number of results in each of the digital and paper formats of the examination in the 38 participating schools. A total of 9,566 results, 46% of which were from the digital examination format, were analysed from 4,681 students.

Table 1. Number of results for Level 1 externally assessed achievement standards in English at the participating schools in the Digital Pilot.

Level 1 English External Achievement Standard		**Total results – digital format**	**Total results – paper format**	**Percent results – digital format**	**Number of participating schools**
90849	Show understanding of specified aspect(s) of studied written text(s), using supporting evidence	1,588	1,796	47%	35
90850	Show understanding of specified aspect(s) of studied visual or oral text(s), using supporting evidence	1,590	1,860	46%	33
90851	Show understanding of significant aspects of unfamiliar written text(s) through close reading, using supporting evidence	1,195	1,537	44%	29
Total		**4,373**	**5,193**	**46%**	**38**

Figure 1 compares the distributions of grades for digital and paper formats for each of the three standards included in the Level 1 English Digital Pilot, and Figure 2 shows the Rasch difficulty parameter estimates for digital and paper formats for the three standards.

For all three standards, the percentages of Achieved grades were higher for the paper format, and the percentages of Excellence grades were higher for the digital format. The percentages for the other grades were variable amongst the three standards. Figure 1 also shows the pairs of differences that were statistically significant ($p<0.5$) – for standard 90850 Achieved grade, where the paper-format students had a higher Achieved grade percentage compared with digital-format students by six percentage points, and the Excellence grade, where the digital-format students had a higher Excellence grade percentage compared with paper-format students by four percentage points.

These results were generally corroborated by the Rasch difficulty parameter estimates shown in Figure 2. For all the three standards, the difficulty parameter estimates associated with attaining grades of Achieved or better were higher for the digital format than for the paper format, indicating that the students found the digital format more difficult than the paper format with respect to gaining credit. The difficulty parameter estimates associated with attaining grades of Merit or better and Excellence were higher for the paper format than the digital format, indicating that the students found the paper format more difficult than the digital format with respect to getting higher grades. However none of the pairs of comparisons were statistically significant ($p<0.5$).

Standard	Result	Percent		Percentage Point Difference	
		Digital	Paper		
90849	N	18%	17%	1%	
	A	41%	46%	-5%	
	M	28%	28%	1%	
	E	12%	10%	3%	
90850	N	23%	21%	2%	
	A	43%	49%	-6%	*
	M	23%	23%	0%	
	E	11%	7%	4%	*
90851	N	16%	17%	-1%	
	A	48%	51%	-3%	
	M	27%	25%	2%	
	E	9%	7%	2%	

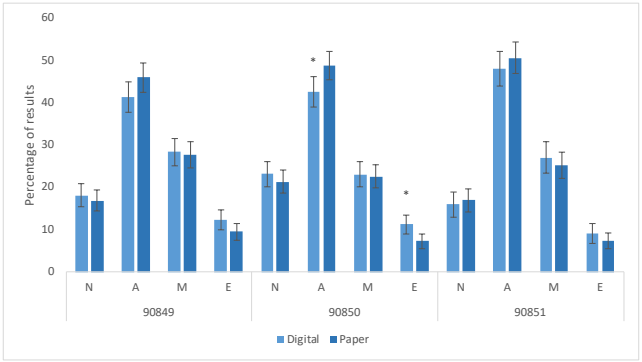

Figure 1. Comparison of digital and paper result distributions for Level 1 externally assessed achievement standards in English at schools participating in the Digital Pilot. Due to rounding, percentages may not add up to 100% and re-calculating the percentage point differences may not be identical to the values displayed in the table. Vertical bars denote 95% confidence intervals, adjusted for multiple comparisons. (* = the difference between digital and paper formats is statistically significant.)

Standard	Result	Difficulty parameter estimate		Difference	
		Digital	Paper		
90849	AME	-1.0	-1.3	0.2	
	ME	0.4	0.5	-0.1	
	E	1.5	1.9	-0.3	
90850	AME	-0.9	-1.1	0.3	
	ME	0.5	0.7	-0.2	
	E	1.5	2.0	-0.5	
90851	AME	-1.2	-1.2	0.1	
	ME	0.4	0.6	-0.2	
	E	1.6	2.1	-0.5	

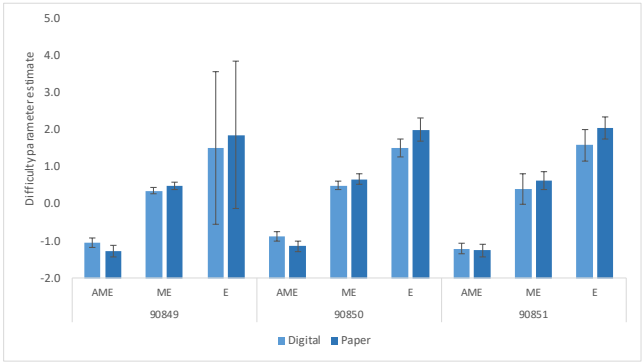

Figure 2. Comparison of Rasch difficulty parameter estimates for attaining grades of Achieved or better (AME), Merit or better (ME), and Excellence (E) for digital and paper formats for each externally assessed Level 1 achievement standard in English. Due to rounding, re-calculating the difficulty parameter estimate differences may not be identical to the values displayed in the table. Vertical bars denote 95% confidence intervals, adjusted for multiple comparisons. (* = the difference between digital and paper formats is statistically significant.)

1.2 Matched data comparison

Analysis described in Appendix A indicated that the findings in the previous section (1.1) were due to the characteristics of the groups of candidates, rather than the characteristics of the mode of external assessment. In order to account for the underlying difference in ability between the digital-format and paper-format students, achievement in internally assessed achievement standards was used to construct matching datasets of digital-format and paper-format students. Any residual differences, after comparing this set of data, could then be attributed to the characteristics of the two examination formats.

Of the 2,020 students in the previous set of analyses who undertook the Level 1 English examination in the digital format, 1,675 students had at least one paper-format student with identical results in Level 1 English internal achievement standards. For this set of matched digital-format students, there was a total of 10,089 paper-format students that were eligible for matching.

A total of 100 resamples were taken from this set of 10,089 paper-format students, with each resample consisting of 1,675 students that had matching internal results to the matched digital-format students. (See Appendix B for the detailed methodology and results.) Figure 3 compares the distribution of grades for the matched digital-format students with the aggregated distribution of results of the matched paper-format student resamples. Figure 4 shows the comparison of the estimated Rasch difficulty parameters for the matched digital-format students with those generated from the matched paper-format student resamples.

For the standards 90849 and 90851, the grade distributions for digital-format students and paper-format students were similar for most grade levels. For 90849, the paper-format students attaining Merit grade is three percentage points higher than that of digital-format students, and the digital-format students attaining Excellence grade is three percentage points higher than that of the paper-format students, but these differences were not statistically significant ($p<0.5$). For 90850, there was a higher percentage of digital-format students who had a Not Achieved grade compared with paper-format students (five percentage points higher), and there was a higher percentage of paper-format students who had a Merit grade compared with digital-format students (six percentage points higher). Both differences were statistically significant ($p<0.5$).

These results were generally corroborated by the Rasch difficulty parameter estimates in Figure 4. As in the previous comparison (Figure 2), the difficulty parameter estimates associated with attaining a grade of Achieved or better were higher for the digital format than for the paper format for all the three standards, which indicated that the students found the digital format more difficult than the paper format with respect to gaining credit. The difficulty parameter estimates associated with attaining Excellence grade was higher for the paper format than the digital format, indicating that the students found the paper format more difficult than the digital format with respect to getting Excellence grade. The only statistically significantly different comparison ($p<0.5$) was for standard 90850, where the difficulty parameter estimate for getting a grade of Achieved or better was higher for the digital format compared with the paper format.

Standard	Result	Percent		Percentage Point Difference	
		Digital	Paper		
90849	N	16%	16%	1%	
	A	42%	42%	0%	
	M	29%	32%	-3%	
	E	13%	10%	3%	
90850	N	22%	17%	5%	*
	A	42%	42%	0%	
	M	24%	30%	-6%	*
	E	12%	10%	2%	
90851	N	15%	14%	1%	
	A	49%	50%	-2%	
	M	27%	27%	1%	
	E	9%	9%	0%	

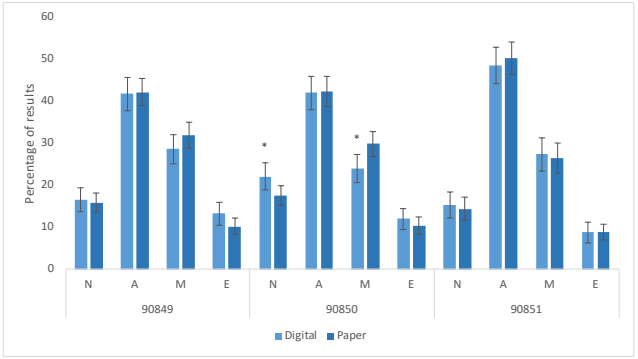

Figure 3. Comparison of result distributions of matched digital-format students and aggregated matched paper-format student resamples (n=100) for Level 1 externally assessed achievement standards in English. Due to rounding, percentages may not add up to 100% and re-calculating the percentage point differences may not be identical to the values displayed in the table. Vertical bars denote 95% confidence intervals, adjusted for multiple comparisons. (* = the difference between digital and paper formats is statistically significant.)

Standard	Result	Difficulty parameter estimate		Difference	
		Digital	Paper		
90849	AME	-1.1	-1.3	0.2	
	ME	0.3	0.3	0.0	
	E	1.5	1.8	-0.3	
90850	AME	-0.9	-1.3	0.4	*
	ME	0.5	0.4	0.1	
	E	1.5	1.7	-0.2	
90851	AME	-1.3	-1.4	0.1	
	ME	0.4	0.6	-0.2	
	E	1.6	1.9	-0.3	

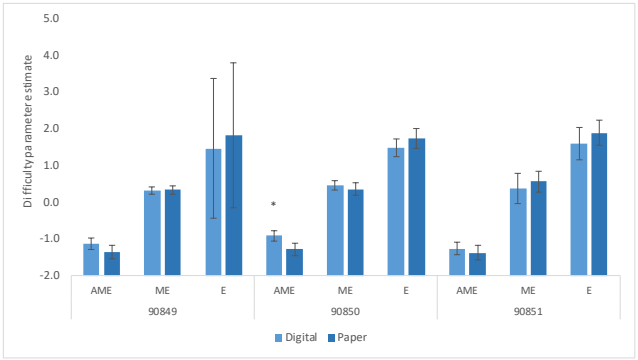

Figure 4. Comparison of Rasch difficulty parameter estimates for attaining grades of Achieved or better (AME), Merit or better (ME), and Excellence (E) for matched digital-format students and aggregated matched paper-format student resamples (n=100) for each externally assessed Level 1 achievement standard in English. Due to rounding, re-calculating the difficulty parameter estimate differences may not be identical to the values displayed in the table. Vertical bars denote 95% confidence intervals, adjusted for multiple comparisons. (* = the difference between digital and paper formats is statistically significant.)

1.3 Conclusion

While some analysis results indicated that the students found the digital format more difficult than the paper format, this finding was not consistent in all the Level 1 English external standards. For standard 90850, there was some evidence of a disadvantage to the digital format with respect to gaining credit, but there was no evidence of a difference for the other standards and grade levels.

2. Level 2 English

2.1 Students in schools participating in Digital Pilot

Table 2 shows the externally-assessed standards in Level 2 English as well as the number of results in each of the digital and paper formats of the examination in the 35 participating schools. A total of 8,058 results, 50% of which were from the digital examination format, were analysed from 4,408 students.

Table 2. Number of results for Level 2 externally assessed achievement standards in English at the participating schools in the Digital Pilot.

Level 2 English External Achievement Standard		**Total results – digital format**	**Total results – paper format**	**Percent results – digital format**	**Number of participating schools**
91098	Analyse specified aspect(s) of studied written text(s), supported by evidence	1,550	1,465	51%	31
91099	Analyse specified aspect(s) of studied visual or oral text(s), supported by evidence	1,292	1,305	50%	30
91100	Analyse significant aspects of unfamiliar written text(s) through close reading, supported by evidence	1,163	1,283	48%	28
Total		**4,005**	**4,053**	**50%**	**35**

Figure 5 compares the distributions of grades for digital and paper formats for each of the three standards included in the Level 2 English Digital Pilot, and Figure 6 shows the Rasch difficulty parameter estimates for the digital and paper formats for the three standards.

For all three standards, the percentages of Not Achieved grades were higher for the paper format, and the percentages of Merit grades were higher for the digital format. The percentages for the other grades were variable amongst the three standards. Figure 5 also shows the pairs of differences that were statistically significant ($p<0.5$). For standard 91099, the paper-format students had a higher Not Achieved grade percentage compared with digital-format students (twelve percentage points higher), and the digital-format students had higher Merit and Excellence grade percentages compared with paper-format students (nine and six percentage points higher respectively). For standard 91100, paper-format students had a higher Not Achieved grade percentage compared with digital-format students (seven percentage points higher), and the digital-format students had a higher Merit grade percentage compared with paper-format students (seven percentage points higher).

These findings were generally corroborated by the Rasch difficulty parameter estimates in Figure 6. All the difficulty parameter estimates were higher for the paper format than for the digital format, indicating that the students found the paper format more difficult than the digital format. Figure 6 also shows which pairs of comparisons were statistically significantly different ($p<0.5$) – attaining a

grade of Achieved or better for standards 91099 and 91100, and attaining a grade of Merit or better for standard 91099.

Standard	Result	Percent		Percentage Point Difference	
		Digital	Paper		
91098	N	22%	26%	-4%	
	A	43%	41%	2%	
	M	25%	24%	2%	
	E	9%	9%	0%	
91099	N	15%	27%	-12%	*
	A	41%	43%	-2%	
	M	30%	21%	9%	*
	E	14%	9%	6%	*
91100	N	17%	24%	-7%	*
	A	45%	46%	-1%	
	M	29%	22%	7%	*
	E	9%	8%	1%	

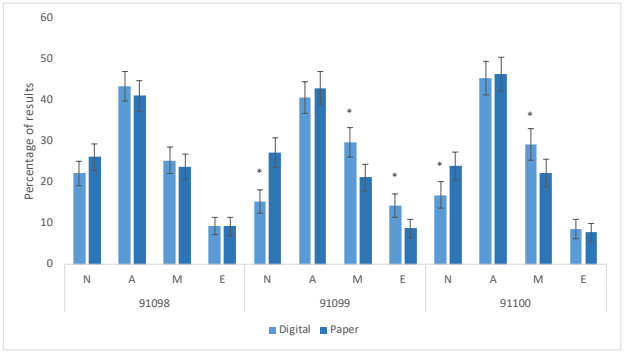

Figure 5. Comparison of digital and paper result distributions for Level 2 externally assessed achievement standards in English at schools participating in the Digital Pilot. Due to rounding, percentages may not add up to 100% and re-calculating the percentage point differences may not be identical to the values displayed in the table. Vertical bars denote 95% confidence intervals, adjusted for multiple comparisons. (* = the difference between digital and paper formats is statistically significant.)

Standard	Result	Difficulty parameter estimate		Difference	
		Digital	Paper		
91098	AME	-1.0	-0.8	-0.1	
	ME	0.6	0.7	-0.1	
	E	1.9	2.0	-0.1	
91099	AME	-1.4	-0.9	-0.5	*
	ME	0.2	0.7	-0.4	*
	E	1.5	1.9	-0.4	
91100	AME	-1.3	-0.9	-0.4	*
	ME	0.4	0.8	-0.4	
	E	1.9	2.1	-0.2	

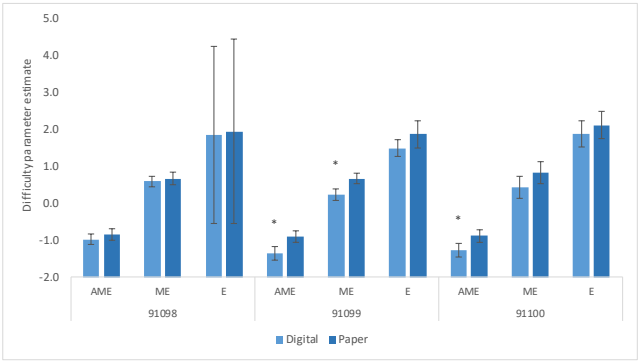

Figure 6. Comparison of Rasch difficulty parameter estimates for attaining grades of Achieved or better (AME), Merit or better (ME), and Excellence (E) for digital and paper formats for each externally assessed Level 2 achievement standard in English. Due to rounding, re-calculating the difficulty parameter estimate differences may not be identical to the values displayed in the table. Vertical bars denote 95% confidence intervals, adjusted for multiple comparisons. (* = the difference between digital and paper formats is statistically significant.)

2.2 Matched data comparison

Analysis described in Appendix A indicated that the findings in the previous section (2.1) were due to the characteristics of the groups of candidates, rather than the characteristics of the mode of external assessment. In order to account for the underlying difference in ability between the digital-format and paper-format students, achievement in internally assessed achievement standards was used to construct matching datasets of digital-format and paper-format students. Any residual differences, after comparing this set of data, could then be attributed to the characteristics of the two examination formats.

Of the 2,085 students in the previous set of analyses who undertook the Level 2 English examination in the digital format, 1,990 students had at least one paper-format student with identical results in Level 2 English internal achievement standards. For this set of matched digital-format students, there was a total of 18,646 paper-format students that were eligible for matching.

A total of 100 resamples were taken from this set of 18,646 paper-format students, with each resample consisting of 1,990 students that had matching internal results to the matched digital-format students. (See Appendix B for the detailed methodology and results.) Figure 7 compares the distribution of grades for the matched digital-format students with the aggregated distribution of results of the matched paper-format student resamples. Figure 8 shows the comparison of the estimated Rasch difficulty parameters for the matched digital-format students with those generated from the matched paper-format student resamples.

The grades distribution for standard 91098 was different from that observed in Figure 5 – the percentages for Not Achieved and Achieved were higher for the digital-format students compared with the paper-format students (by four and three percentage points respectively), and the percentages for Merit and Excellence were higher for the paper-format students compared with the digital-format students (both by four percentage points). Also, the percentage differences between the two formats were statistically significantly different for the grades of Not Achieved and Excellence ($p<0.5$). For the other standards, the differences between the two examination formats were not statistically significant across all grade levels ($p<0.5$).

These findings were corroborated by the Rasch difficulty parameter estimates in Figure 8. All the difficulty parameters for 91098 were higher for the digital format compared with the paper format, which indicated that for this standard, the students found the digital format more difficult than the paper format. Also, for all the standards, the difficulty parameter associated with attaining Excellence grade was higher for the digital format compared with the paper format, which indicated that the students found the digital format more difficult than the paper format with respect to gaining an Excellence grade. However the only statistically significantly different comparison was attaining a grade of Merit or better for 91098.

Standard	Result	Percent		Percentage Point Difference	
		Digital	Paper		
91098	N	22%	17%	4%	*
	A	43%	40%	3%	
	M	26%	30%	-4%	
	E	9%	13%	-4%	*
91099	N	15%	17%	-2%	
	A	40%	41%	-1%	
	M	30%	27%	3%	
	E	15%	15%	0%	
91100	N	17%	16%	0%	
	A	45%	45%	0%	
	M	30%	27%	3%	
	E	9%	11%	-3%	

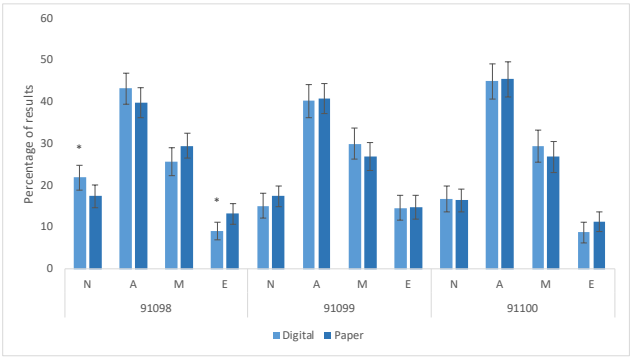

Figure 7. Comparison of result distributions of matched digital-format students and aggregated matched paper-format student resamples (n=100) for Level 2 externally assessed achievement standards in English. Due to rounding, percentages may not add up to 100% and re-calculating the percentage point differences may not be identical to the values displayed in the table. Vertical bars denote 95% confidence intervals, adjusted for multiple comparisons. (* = the difference between digital and paper formats is statistically significant.)

Standard	Result	Difficulty parameter estimate		Difference	
		Digital	Paper		
91098	AME	-1.0	-1.3	0.3	
	ME	0.6	0.3	0.3	*
	E	1.9	1.6	0.3	
91099	AME	-1.4	-1.3	-0.1	
	ME	0.2	0.3	-0.1	
	E	1.5	1.4	0.1	
91100	AME	-1.3	-1.3	0.0	
	ME	0.4	0.5	-0.1	
	E	1.9	1.8	0.1	

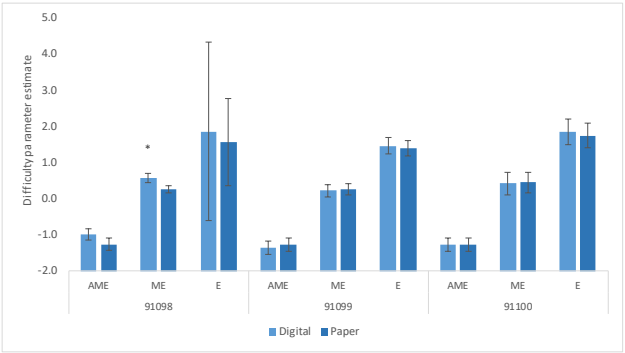

Figure 8. Comparison of Rasch difficulty parameter estimates for attaining grades of Achieved or better (AME), Merit or better (ME), and Excellence (E) for matched digital-format students and aggregated matched paper-format student resamples (n=100) for each externally assessed Level 2 achievement standard in English. Due to rounding, re-calculating the difficulty parameter estimate differences may not be identical to the values displayed in the table. Vertical bars denote 95% confidence intervals, adjusted for multiple comparisons. (* = the difference between digital and paper formats is statistically significant.)

2.3 Conclusion

While some analysis results indicated that the students found the digital format more difficult than the paper format, this finding was not consistent in all the Level 2 English external standards. For standard 91098, there was some evidence of a disadvantage to the digital format with respect to gaining a grade of Merit or better, but there was no evidence of a difference for the other standards and grade levels.

Appendix

A. Analysis of predictive relationship between internal and external assessment

Since students participating in the digital assessment were self-selecting, there was no basis to assume that the digital-format and paper-format students were equal in ability, or that they ought to have attained the same distribution of results for their external assessments. Therefore results from Sections 1.1 for Level 1 English and 2.1 for Level 2 English might be explained by the mode of external assessment (digital vs on paper) as well as difference in underlying ability between the two groups of students.

To further investigate the differences discussed in Sections 1.1 and 2.1, a set of linear regression analyses was conducted to assess the equivalence of digital and paper assessment mediums in terms of the extent to which the level of performance in each format predicts a given level of performance in the internally assessed achievement standards. If such an equivalence was established, it could be concluded that the observed differences in Sections 1.1 for Level 1 English and 2.1 for Level 2 English were due to difference between the two groups of students, rather than difference between the characteristics of the two modes of external assessment. Moreover, performance in internally-assessed standards can be used in subsequent analyses to account for underlying difference in ability between the two groups of students.

For each of Level 1 and Level 2 English, a set of Rasch analysis was carried out on all internally assessed results of participating candidates – both those who completed all the external assessments digitally and those who completed all of them on paper. Like the analyses of the external assessment formats, these analyses 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 then used to model the predictive relationship between external and internal assessment performance for the digital-format students, and another to model the predictive relationship between external and internal assessment performance for the paper-format students. Figure 9 depicts the two scatterplots, with regression lines, for Level 1 English, showing the relationship between external and internal assessments for each of the digital-format and paper-format students. Figure 10 shows the corresponding graphs for Level 2 English.

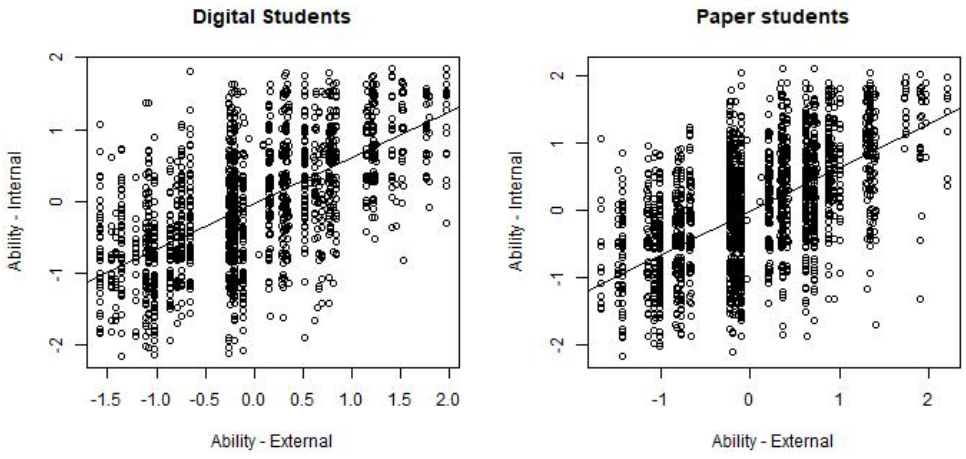

Figure 9. Scatterplots with regression lines showing the relationships between Rasch ability parameters estimated for Level 1 English external and internal assessments. (left: digital-format students; right: paper-format students)

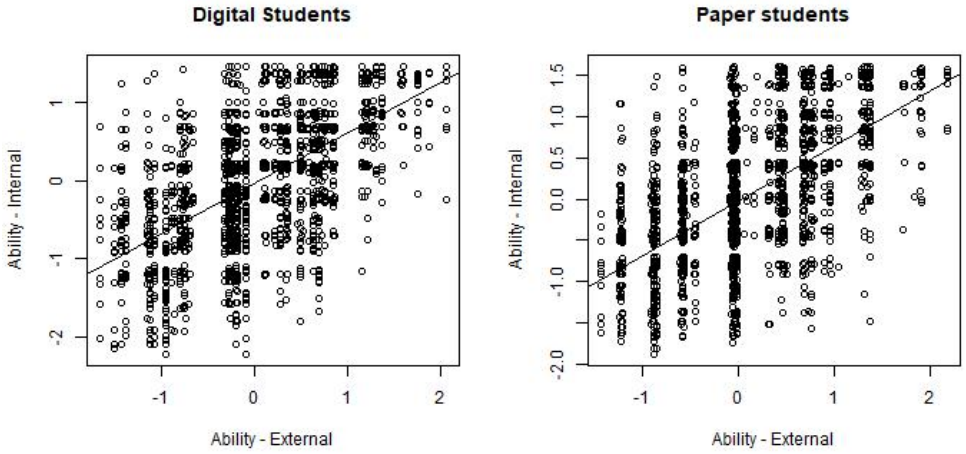

Figure 10. Scatterplots with regression lines showing the relationships between Rasch ability parameters estimated for Level 2 English external and internal assessments. (left: digital-format students; right: paper-format students)

Table 3 shows the constant and slope parameters estimated by the regression models for Level 1 and Level 2 English. Comparison of these parameters would allow for a statistical comparison of the equivalence of the predictive relationships between externals and internals achievement. The set of regression models did not differ significantly in their parameter estimates, suggesting that a candidate with a given level of attainment in internal assessment is predicted to achieve the same level of attainment in external assessment, irrespective of whether the external assessment is conducted in a digital or paper-based format. This conclusion was the same as the psychometric analysis of 2017 results for Level 1 English[4].

Table 3. Parameter estimates of the pairs of regression models from Figure 9 (Level 1 English) and Figure 10 (Level 2 English). Due to rounding, re-calculating the estimate raw differences may not be identical to the values displayed in the table. The differences between the constant and slope parameters are inside the 95% confidence intervals. None of the pairs of differences are statistically significantly different.

Examination	Parameter	Estimates and 95% Confidence Intervals		Estimate Raw Difference
		Digital	Paper	
Level 1 English	intercept	-0.02 (-0.04, 0.01)	-0.01 (-0.03, 0.02)	-0.01
	slope	0.64 (0.60, 0.67)	0.64 (0.61, 0.68)	-0.01
Level 2 English	intercept	-0.02 (-0.05, 0.00)	-0.02 (-0.04, 0.01)	-0.01
	slope	0.65 (0.61, 0.68)	0.66 (0.63, 0.70)	-0.02

This finding suggests that it is the characteristic of the groups of candidates, rather than characteristics of the mode of external assessment, that would explain the differences in grade distributions and Rasch difficulty parameters in Sections 1.1 for Level 1 English and 2.1 for Level 2 English.

This then forms the rationale behind the matched data comparison described in Sections 1.2 for Level 1 English and 2.2 for Level 2 English, where matched digital-format and paper-format students were

[4] For more information, refer to https://www.nzqa.govt.nz/assets/About-us/Future-State/2017-trials-and-pilots/Psychometricanalysesreport.pdf

compared by matching on internally assessed achievement standards. Any residual differences could then be attributed to the characteristics of the two examination formats.

B. Matched data analysis methodology detail and results

The analyses performed in sections 1.2 for Level 1 English and 2.2 for Level 2 English are described in detail below.

A. Identify matching sets of students
   1. Identify all digital-format students with at least one paper-format student with identical results in internal achievement standards at the same level.
   2. Identify all paper-format students that are eligible for matching to students identified in Step 1 i.e. paper-format students with at least one digital-format student with identical results in internal achievement standards at the same level.

B. Generate 100 resamples and perform analysis
   3. For each digital-format student identified in Step 1:
      a. Identify all paper-format students with identical results in internal achievement standards at the same level.
      b. From the list of paper-format students identified in Step 3a, randomly sample one student.
   4. Calculate the grade distribution for externally assessed standards of matched paper-format students from Step 3.
   5. Conduct Rasch analysis to estimate the difficulty parameters for externally assessed standards of matched paper-format students from Step 3.
   6. Perform Steps 3 to 5 100 times.

C. Compile summary of results
   7. Calculate the average of the 100 percentages, for each external standard-grade combination, generated from Step 6 (see Figure 11 for Level 1 English results and Figure 13 for Level 2 English results).
   8. Calculate the mode (most commonly occurring value) of the 100 difficulty parameter estimates, for each external standard-difficulty combination, generated from Step 6 (see Figure 12 for Level 1 English results and Figure 14 for Level 2 English results).
   9. Compare the results of Steps 7 and 8 with the corresponding values from matched digital-format students.

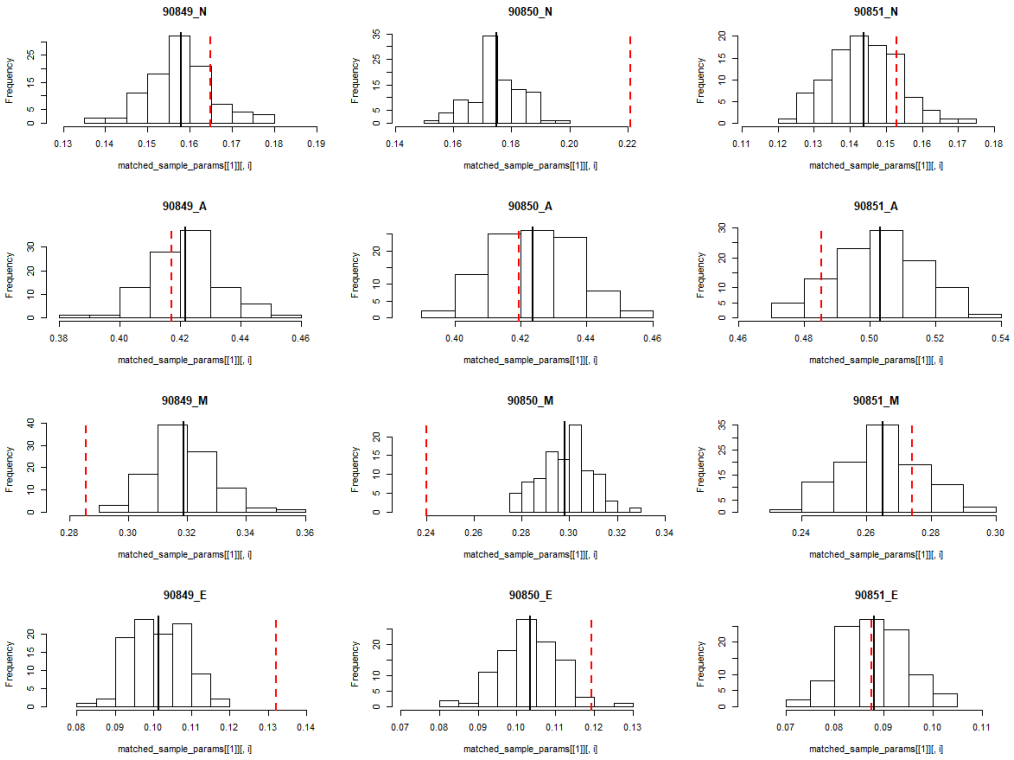

Figure 11. Grade distributions from Level 1 English externally-assessed achievement standards of matched paper-format student resamples (n=100). Vertical solid black lines denote the average percentage. Vertical dotted red lines denote the digital-format student percentage for the corresponding external standard and grade level.

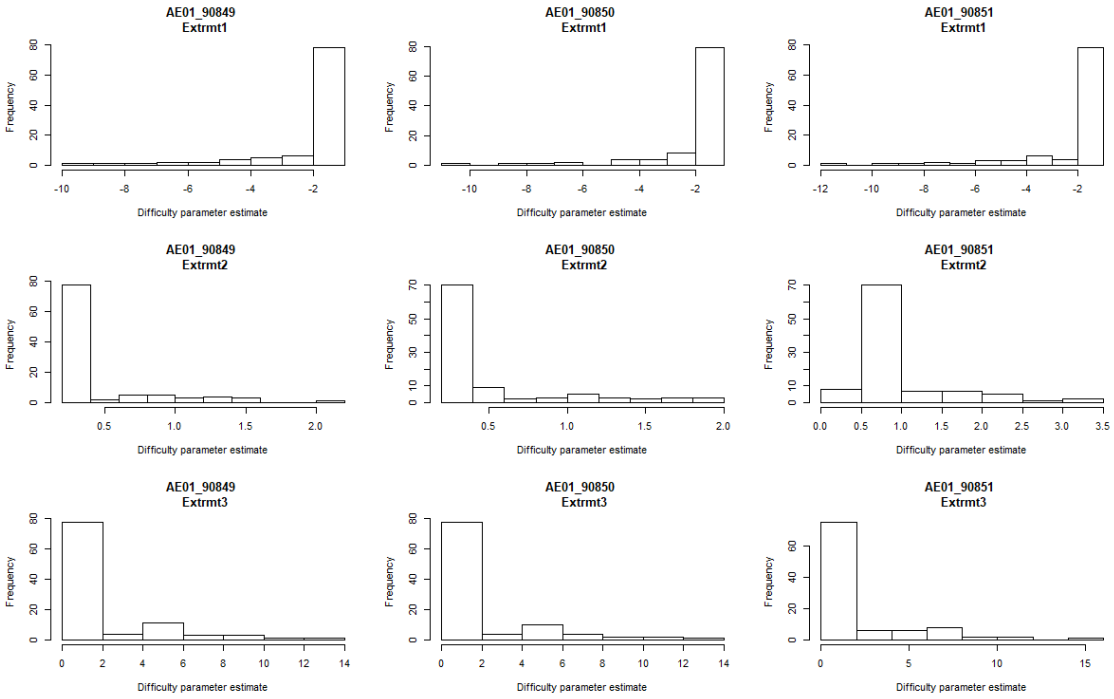

Figure 12. Difficulty parameter estimates of Level 1 English externally-assessed achievement standards from Rasch analysis of matched paper-format student resamples (n=100).

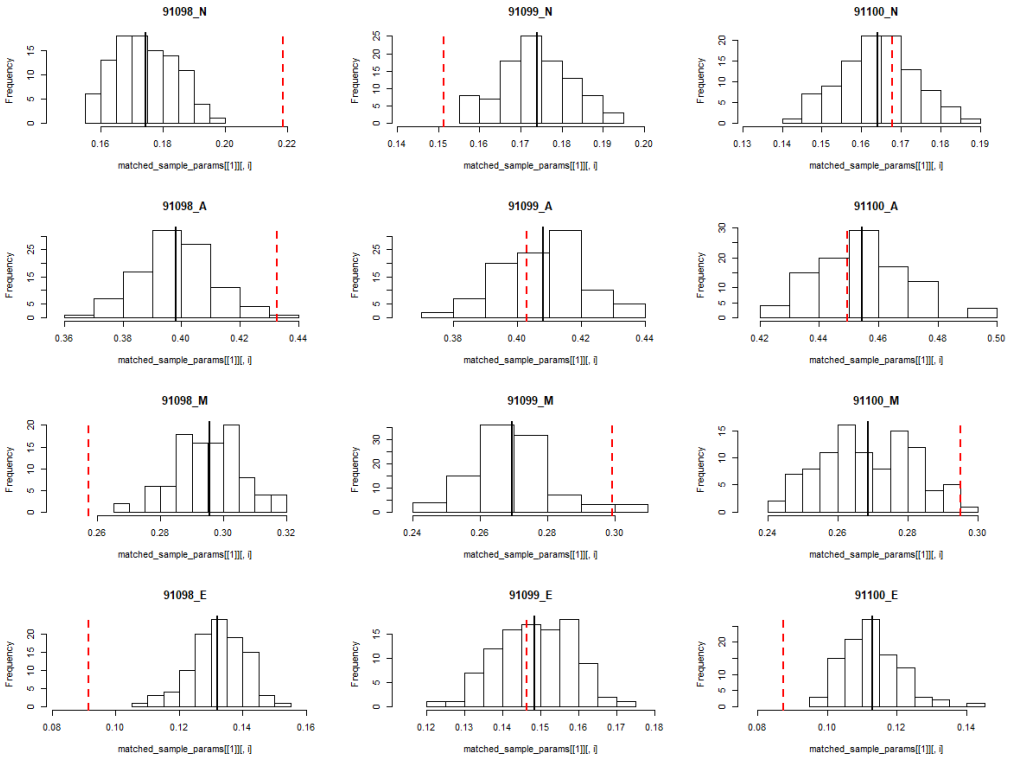

Figure 13. Grade distributions from Level 2 English externally-assessed achievement standards of matched paper-format student resamples (n=100). Vertical solid black lines denote the average percentage. Vertical dotted red lines denote the digital-format student percentage for the corresponding external standard and grade level.

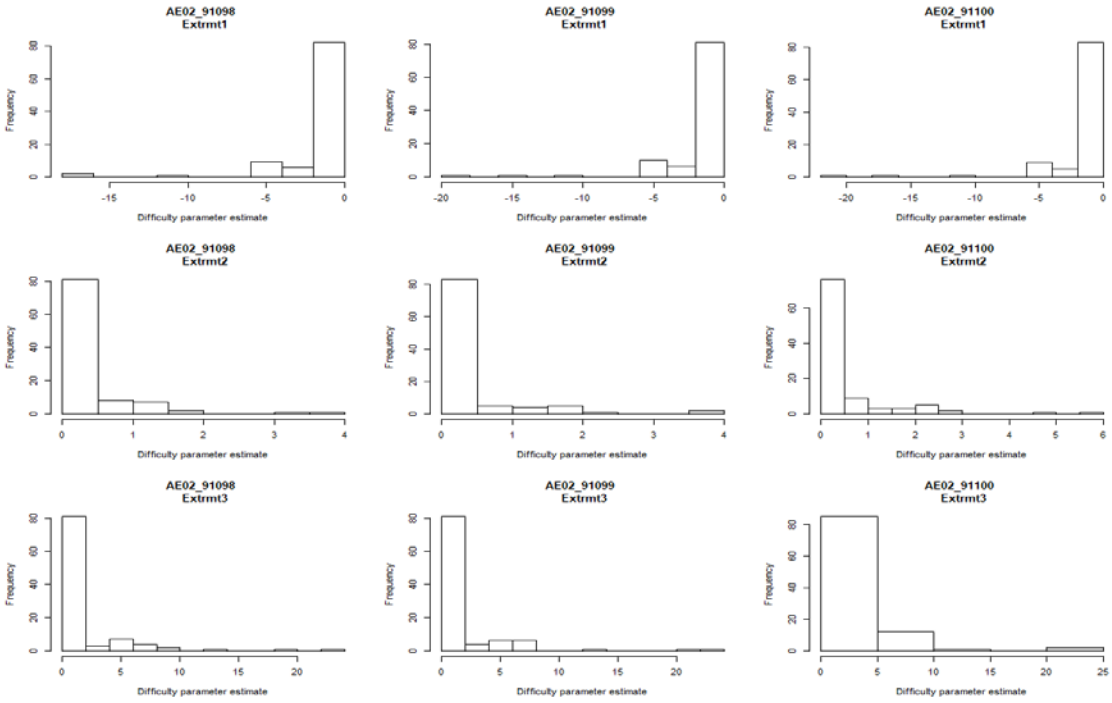

Figure 14. Difficulty parameter estimates of Level 2 English externally-assessed achievement standards from Rasch analysis of matched paper-format student resamples (n=100).

C. Level 1 English: Number of results in each grade category for each participating school, disaggregated by assessment format

90849	Digital					Paper				
School Index #	%N	%A	%M	%E	n	%N	%A	%M	%E	n
1	60	40	0	0	5	0	0	0	0	0
2	0	0	0	100	1	32	38	21	10	63
3	15	23	15	46	13	16	39	34	10	119
4	32	54	11	3	71	16	47	26	11	19
5	33	33	33	0	3	16	48	30	7	149
6	41	49	8	3	37	53	33	13	0	15
7	75	23	2	0	56	40	40	20	0	10
8	16	55	18	10	49	15	62	8	15	13
9	25	33	33	8	12	16	34	39	11	44
10	24	66	10	0	29	0	0	0	0	0
11	30	39	21	9	33	24	39	29	8	75
12	10	81	10	0	21	9	55	26	9	121
13	0	0	0	0	0	24	51	19	6	68
14	33	52	7	9	46	17	50	33	0	6
15	17	43	32	8	75	50	42	0	8	12
16	0	0	0	0	0	18	37	32	13	87
17	7	36	37	20	292	9	53	22	16	32
18	12	47	31	9	32	12	45	31	12	74
19	0	100	0	0	1	11	53	31	5	55
20	5	33	43	19	21	6	52	29	13	31
21	17	56	20	7	41	0	0	0	0	0
22	5	47	21	26	19	9	44	34	12	99
23	14	43	29	14	21	0	0	0	100	1
24	27	36	31	6	78	29	38	21	12	34
25	19	50	26	5	122	16	46	28	9	95
26	40	40	0	20	5	16	59	23	3	227
27	0	0	0	0	0	0	0	0	0	0
28	45	41	14	0	22	50	50	0	0	6
29	2	21	47	30	47	0	38	38	24	45
30	73	23	5	0	22	100	0	0	0	6
31	10	34	37	19	59	9	24	37	30	46
32	20	80	0	0	5	15	68	12	6	34
33	3	44	25	28	36	11	38	32	19	47
34	29	61	11	0	28	23	14	46	17	35
35	2	23	52	23	48	0	0	100	0	2
36	13	32	35	19	31	0	0	0	0	0
37	6	40	42	12	202	14	51	28	7	43
38	40	60	0	0	5	24	52	20	4	83
Total	**18**	**41**	**28**	**12**	**1,588**	**17**	**46**	**28**	**10**	**1,796**

90850	Digital					Paper				
School Index #	%N	%A	%M	%E	n	%N	%A	%M	%E	n
1	36	40	24	0	45	0	0	0	0	0
2	0	0	0	0	0	44	33	18	5	39
3	7	57	21	14	14	24	48	21	8	102
4	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	23	61	16	0	82
6	24	52	22	2	50	65	23	8	4	26
7	75	23	2	0	61	50	38	12	0	8
8	15	58	19	8	62	8	54	15	23	13
9	40	33	13	13	15	13	51	11	26	47
10	31	56	11	2	55	0	100	0	0	1
11	100	0	0	0	3	47	49	4	0	51
12	17	75	8	0	24	16	76	8	0	102
13	44	56	0	0	16	26	47	21	6	89
14	40	50	4	6	50	40	20	40	0	5
15	23	53	16	8	90	55	18	27	0	11
16	75	25	0	0	20	15	52	26	7	113
17	14	38	28	20	302	16	50	28	6	32
18	12	34	34	19	32	11	47	37	5	73
19	27	52	19	2	48	16	42	27	15	123
20	4	43	22	30	23	16	47	31	6	32
21	29	46	11	14	28	0	100	0	0	1
22	10	40	35	15	20	15	38	43	5	103
23	10	40	35	15	20	0	100	0	0	1
24	32	42	23	3	65	35	32	24	9	34
25	19	44	32	5	111	19	38	30	13	79
26	50	33	0	17	6	22	59	18	1	282
27	0	0	0	0	0	20	60	20	0	5
28	61	22	11	6	18	0	100	0	0	1
29	7	20	49	24	45	2	34	41	23	44
30	59	33	7	0	27	70	30	0	0	10
31	0	0	0	0	0	16	37	34	13	89
32	20	80	0	0	5	27	49	19	5	37
33	24	29	24	24	21	22	41	28	9	32
34	18	71	11	0	28	21	44	16	19	77
35	0	19	35	46	52	0	0	100	0	2
36	4	56	33	7	27	0	0	0	0	0
37	12	43	33	12	187	10	45	28	17	29
38	50	40	10	0	20	24	59	15	2	85
Total	**23**	**43**	**23**	**11**	**1,590**	**21**	**49**	**23**	**7**	**1,860**

90851	Digital					Paper				
School Index #	%N	%A	%M	%E	n	%N	%A	%M	%E	n
1	26	40	31	3	72	14	86	0	0	7
2	0	0	0	0	0	16	47	27	11	64
3	15	46	23	15	13	15	54	28	4	112
4	15	51	28	6	72	11	63	16	11	19
5	0	25	25	50	4	6	26	49	18	65
6	0	0	0	0	0	0	0	0	0	0
7	79	19	2	0	52	100	0	0	0	4
8	13	56	23	8	62	15	38	38	8	13
9	0	67	22	11	18	13	48	29	10	52
10	0	46	46	8	13	0	0	100	0	1
11	9	61	21	9	33	16	64	16	4	80
12	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	16	59	21	4	126
14	20	60	12	8	40	12	38	38	12	8
15	11	53	33	3	75	14	71	14	0	14
16	0	0	0	0	0	0	0	0	0	0
17	19	53	19	10	59	25	50	17	8	12
18	10	52	26	13	31	21	47	25	8	73
19	0	0	0	0	0	0	31	38	31	29
20	5	43	29	24	21	26	44	24	6	34
21	24	50	24	3	34	0	100	0	0	1
22	10	30	35	25	20	15	48	29	8	100
23	14	52	33	0	21	0	100	0	0	1
24	15	54	26	5	80	18	53	24	5	38
25	25	58	16	1	100	26	52	21	2	66
26	0	33	33	33	3	21	64	13	2	270
27	0	50	50	0	2	0	22	41	38	32
28	0	80	20	0	5	0	0	0	0	0
29	0	26	39	35	46	0	34	39	26	38
30	0	0	0	0	0	0	0	0	0	0
31	0	0	0	0	0	100	0	0	0	1
32	0	100	0	0	5	22	38	41	0	37
33	9	28	49	15	47	17	32	42	10	60
34	12	84	3	0	32	22	33	33	11	9
35	0	0	43	57	7	0	0	0	0	0
36	0	0	0	0	0	0	0	0	0	0
37	5	44	38	13	208	16	47	29	8	49
38	30	40	25	5	20	20	53	22	4	122
Total	**16**	**48**	**27**	**9**	**1,195**	**17**	**51**	**25**	**7**	**1,537**

D. Level 2 English: Number of results in each grade category for each participating school, disaggregated by assessment format

91098	Digital					Paper				
School Index #	%N	%A	%M	%E	n	%N	%A	%M	%E	n
1	19	49	24	9	70	43	40	14	4	168
2	21	39	24	15	33	36	36	18	9	11
3	67	33	0	0	6	23	39	26	12	104
4	13	28	36	23	39	19	40	32	9	85
5	31	53	7	9	45	48	41	10	0	29
6	0	0	0	0	0	0	0	50	50	2
7	7	33	44	15	27	20	62	13	4	45
8	0	62	25	12	8	0	0	57	43	7
9	54	41	5	0	37	31	46	21	3	39
10	0	0	0	0	0	37	30	27	6	109
11	22	44	25	9	247	48	27	18	6	33
12	25	50	25	0	12	30	52	9	9	23
13	36	29	29	7	14	0	0	0	0	0
14	0	25	50	25	4	13	37	28	22	95
15	0	0	0	0	0	0	0	0	0	0
16	0	50	33	17	12	31	44	19	5	118
17	57	29	14	0	7	0	0	0	0	0
18	7	47	35	12	75	33	0	50	17	6
19	15	28	32	26	47	0	0	0	0	0
20	10	50	30	10	10	13	49	30	7	69
21	28	38	23	10	212	12	38	38	12	16
22	0	0	0	0	0	0	0	0	0	0
23	20	44	29	8	268	18	42	28	12	139
24	39	56	6	0	18	21	55	21	3	58
25	20	80	0	0	5	33	0	33	33	3
26	60	20	20	0	5	34	50	16	0	44
27	100	0	0	0	2	70	30	0	0	10
28	32	52	14	2	91	28	38	26	8	50
29	50	33	17	0	12	67	0	33	0	3
30	29	71	0	0	7	50	33	17	0	6
31	23	33	35	9	43	23	42	19	15	26
32	4	44	40	13	78	6	45	18	30	33
33	16	53	32	0	19	8	58	26	8	50
34	17	60	21	2	58	5	19	57	19	21
35	26	41	18	15	39	22	33	30	14	63
Total	**22**	**43**	**25**	**9**	**1,550**	**26**	**41**	**24**	**9**	**1,465**

91099	Digital					Paper				
School Index #	%N	%A	%M	%E	n	%N	%A	%M	%E	n
1	12	55	25	8	73	35	49	13	3	205
2	19	47	19	14	57	20	50	10	20	20
3	25	50	25	0	4	37	23	29	11	35
4	12	50	25	12	8	36	36	21	7	14
5	22	52	18	8	50	41	47	6	6	34
6	0	0	0	0	0	33	50	0	17	6
7	21	5	53	21	19	24	50	17	9	66
8	9	45	36	9	11	0	0	67	33	3
9	100	0	0	0	1	27	53	15	5	60
10	100	0	0	0	4	38	39	16	6	159
11	12	44	30	14	238	32	45	15	8	40
12	9	64	27	0	11	11	52	26	11	27
13	36	64	0	0	11	0	0	100	0	2
14	0	0	67	33	3	13	40	31	16	109
15	0	50	50	0	2	29	47	18	6	17
16	0	58	25	17	12	23	40	29	8	87
17	9	73	18	0	11	0	0	0	0	0
18	3	32	42	23	62	17	33	33	17	6
19	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	42	47	0	11	19
21	19	40	26	15	187	22	33	28	17	18
22	16	53	21	11	19	100	0	0	0	1
23	8	38	34	19	229	17	44	25	14	102
24	14	64	23	0	22	23	42	29	6	48
25	19	42	31	8	26	20	80	0	0	5
26	0	0	0	0	0	0	0	0	0	0
27	80	20	0	0	5	73	27	0	0	11
28	33	33	24	11	46	50	39	11	0	28
29	79	10	10	0	29	100	0	0	0	2
30	33	44	22	0	9	55	36	9	0	11
31	5	42	42	11	19	12	42	38	8	24
32	1	32	40	27	81	6	39	33	21	33
33	0	0	0	0	0	0	0	0	0	0
34	100	0	0	0	1	23	40	26	11	53
35	12	24	48	17	42	13	38	37	12	60
Total	**15**	**41**	**30**	**14**	**1,292**	**27**	**43**	**21**	**9**	**1,305**

91100	Digital					Paper				
School Index #	%N	%A	%M	%E	n	%N	%A	%M	%E	n
1	25	45	26	4	73	39	46	14	1	145
2	23	44	20	13	79	22	48	13	17	23
3	0	0	0	0	0	2	36	44	18	50
4	0	50	50	0	2	0	50	50	0	4
5	0	0	0	0	0	0	0	0	0	0
6	0	67	0	33	3	10	32	42	16	31
7	13	43	39	4	23	17	43	30	9	53
8	0	0	0	0	0	0	0	0	0	0
9	50	50	0	0	2	26	48	20	6	81
10	50	50	0	0	4	39	40	17	3	179
11	37	26	26	11	19	0	50	25	25	4
12	17	42	42	0	12	25	57	11	7	28
13	18	64	13	5	39	0	0	50	50	2
14	0	0	0	100	1	18	48	26	9	105
15	0	100	0	0	1	18	71	12	0	17
16	0	50	50	0	2	25	48	18	9	56
17	0	0	0	0	0	0	0	0	0	0
18	50	0	0	50	2	0	0	0	0	0
19	2	57	33	9	58	0	0	0	0	0
20	18	64	18	0	11	28	50	17	5	76
21	21	51	21	6	201	20	60	13	7	15
22	16	47	37	0	19	0	100	0	0	1
23	10	46	35	9	244	13	41	34	11	87
24	0	0	0	0	0	0	0	0	0	0
25	0	100	0	0	1	0	100	0	0	1
26	29	43	29	0	7	35	54	9	2	57
27	0	0	0	0	0	0	0	0	0	0
28	20	49	23	8	91	24	48	24	4	50
29	45	36	0	18	11	100	0	0	0	1
30	0	0	0	0	0	50	50	0	0	2
31	12	29	37	22	51	12	46	25	17	48
32	1	19	62	18	79	0	48	19	32	31
33	5	53	37	5	19	14	55	25	6	51
34	29	50	19	3	70	12	25	50	12	8
35	28	38	26	8	39	19	48	25	8	77
Total	**17**	**45**	**29**	**9**	**1,163**	**24**	**46**	**22**	**8**	**1,283**