

Technical Report # 48

Scoring of a Nationally Representative Sample of Student Writing at Level 5—6 of the New Zealand English Curriculum: Use and Refinement of the asTTle Progress Indicators & Tasks

Abstract: This report overviews the procedures and results of a scoring panel of some 7000 writing scripts completed by students in Years 8 to 12 in response to 35 writing prompts. The panel consisted of some 50 Auckland-based secondary school teachers who were trained in the use of the asTTle progress indicators for 7 different purposes (i.e., analyse, persuade, instruct, describe, narrate, explain, recount) across the range of curriculum levels 2 to 6. Each script was scored for seven dimensions (i.e., audience, content, structure, language, grammar, punctuation, and spelling). Feedback from panellists in the process of using the rubrics led to the revisions of the indicators. Evidence for the reliability of scoring is reported using consensus, consistency, and measurement approaches. High levels of reliability (consensus adjacent agreement > 67%; consistency $\alpha > .80$; measurement $\phi > .80$) were demonstrated giving users confidence in both the scoring rubrics supplied with asTTle and the norms embedded in the asTTle application.



Submitted by the Assessment Tools for Teaching and Learning team,

Auckland UniServices Ltd

University of Auckland

August 2004

Scoring of a Nationally Representative Sample of Student Writing at Level 5—6 of the New Zealand English Curriculum: Use and Refinement of the asTTle Progress Indicators & Tasks

asTTle is funded by the Ministry of Education to Auckland UniServices at the University of Auckland to research and develop an assessment application for Reading, Writing, Mathematics, Pānui, Pāngarau, and Tuhituhi for Years 5-11 (Levels 2-6) for New Zealand schools. We acknowledge this funding, and thank the Ministry of Education for their continued assistance in the development of this project.

This report overviews the procedures and results of a scoring panel of some 7000 writing scripts completed by students in Years 8 to 12 in response to 35 writing prompts. The panel consisted of some 50 Auckland-based secondary school teachers who were trained in the use of the asTTle progress indicators for 7 different purposes (i.e., analyse, persuade, instruct, describe, narrate, explain, recount) across the range of curriculum levels 2 to 6. Each script was scored for seven dimensions (i.e., audience, content, structure, language, grammar, punctuation, and spelling). Feedback from panellists in the process of using the rubrics led to the revisions of the indicators. Evidence for the reliability of scoring is reported using consensus, consistency, and measurement approaches. High levels of reliability (consensus adjacent agreement > 67%; consistency $\alpha > .80$; measurement $\phi > .80$) were demonstrated giving users confidence in both the scoring rubrics supplied with asTTle and the norms embedded in the asTTle application.

I would like to thank all the students and teachers across the country for completing the writing prompts. I especially acknowledge the Auckland secondary school teachers who spent four days of their non-contact time learning to use the asTTle progress indicators and score the scripts.

Dr Gavin Brown managed the process, facilitated the panel, conducted the reliability studies, facilitated the rubric revisions, and wrote this report. Kerry Sussex, Amy Hodgson, and Raquel Guttenbeil provided logistic support throughout the panel ensuring that the many scripts stayed organised and ordered. Earl Irving devised and trained the panellists in the reliability monitoring processes. Kerry summarised the teacher evaluation material.



John Hattie
Project Director, asTTle
August, 2004

The bibliographic citation for this report is:

Brown, G.T.L., Irving, S.E. & Sussex, K. (2004, August). *Scoring of a nationally representative sample of student writing at Level 5—6 of the New Zealand English curriculum: Use and refinement of the asTTle progress indicators & tasks*. asTTle Technical Report 48, University of Auckland/Ministry of Education.

Table of Contents

Materials.....	3
Participants.....	5
Training.....	5
Accuracy Monitoring.....	6
Accuracy Estimation with the Daily Common Script.....	6
Workshop Evaluation.....	11
Physical Environment.....	11
Workshop activities.....	12
Anticipated benefits of asTTle.....	13
Discussion.....	13
References.....	15
Appendix 1. Progress Indicators Levels 2—6.....	16
Appendix 2. Training Materials for the Scoring of Writing Levels 2—6.....	33

Consistent with the approach taken by asTTle in standardising and calibrating materials for levels 2—4 in writing (Glasswell & Brown, 2003), a large-scale marking panel was convened to score a nationally representative sample of writing completed by students in Years 8 to 12 during Term 1, 2004. As reported earlier (Brown, Pipe, & Higginson, 2003), 35 writing tasks and draft progress indicators for seven purposes of writing (i.e., persuade, instruct, narrate, describe, explain, recount, and analyse) were piloted in late 2003. 7337 scripts from 56 schools were returned by schools on behalf of the project. Those rubrics had been revised by the asTTle team and Dr Judy Parr, University of Auckland and presented to panel members as confidential drafts for which their feedback was sought.

The goals of the panel were to (a) complete the marking of 7,000 scripts, (b) improve the progress indicators, and (c) to refine the writing prompts. This report analyses the reliability of the scoring of the writing and provides the updated scoring rubrics (Appendix 1). Note that this report concerns the analytic scoring, (i.e., seven scores per script) of timed impromptu tests (Weigle, 2002) that were obtained through a national large-scale testing program.

Materials

In order to create a manageable yet generalisable data set, it was decided to mark 20 scripts from each school participating in each task. Thus, approximately 240 scripts (12 schools by 20 students) for each task, and about 1200 scripts (5 tasks per purpose) were scored. Table 1 shows the number of scripts scored during the panel by year and purpose. Note that since the

analyse purpose was created for use with secondary students only, no students below Year 11 were assigned the tasks and no criteria for levels below Level 4 were drafted or used.

Table 1.

Scripts Scored by Purpose and Year.

Purpose	Y8	Y9	Y10	Y11	Y12	Missing	Total
Persuade	51	349	171	202	241	10	1024
Instruct	67	351	181	155	272	7	1033
Narrate	49	338	221	236	226	12	1082
Describe	64	371	176	166	241	13	1031
Explain	74	339	195	203	277	5	1093
Recount	77	293	209	171	245	7	1002
Analyse	49	344	206	227	239	7	1072
Total	431	2385	1359	1360	1741	61	7337

The scoring rubrics or progress indicators used in this workshop consisted of a set of ordered descriptors for seven scoring categories (i.e., audience awareness and purpose, content, structure or organisation, language resources, grammar, punctuation, and spelling). For each category, there were five Curriculum Level Proficient descriptors (i.e., one for each of Level 2 to 6); except for analyse which had descriptors only for Levels 4 to 6. Markers identified the Curriculum Level descriptor that best fitted the quality of the writing being marked and then judged whether augmentation of the Level grade was necessary; in other words, the writing was assigned to a Level and then subsequently judged as appropriate (Proficient), weak (Basic), or strong (Advanced) within that level. The process of augmentation (Penny, Johnson, & Gordon, 2000), created a scale with a possible 17 score points (Below 2 Basic to Above 6 Advanced) for each score variable within each purpose, except for the Analyse purpose which only had a score range of 11 points (Below 4 Basic to Above 6 Advanced). Scales of this length are considered an important factor in raising the reliability of scoring (Coffman, 1971).

Participants

Markers were invited to participate via an invitation sent to the HOD English at all Auckland secondary schools. All self-nominating teachers provided details of their teaching experience and any previous official qualifications marking experience, in order to assist selection of the most appropriate candidates for marking. Selection criteria ensured that teachers were registered, had taught at all age groups across curriculum levels 4 to 6 and years 9 to 12 and had been teaching in New Zealand at least two years. This ensured that markers had a thorough knowledge of the curriculum, standards, and students. 50 teachers were selected, of whom 48 attended the panel. The teachers were systematically assigned into panels of 5 to 7 markers to ensure that each panel contained a diversity length of teaching experience and type of school currently employed in. Wherever possible, each panel was lead by a teacher who had previous experience with the design or use of asTTle™ writing rubrics (e.g., teachers who had contributed to asTTle™ Technical Reports 29 and 45). Each panel was assigned to marking one purpose only.

Training

Prior to the marking panel convening, markers were provided with a guide that outlined the scoring, marking and check marking procedures and asked to familiarise themselves with these procedures (see appendix 2). The first morning was devoted to training, and this occurred in two parts. The first part included a demonstration of the asTTle application, with a focus on the assessment and reporting of writing, and a review of the scoring and marking procedures designed to ensure that all markers used the same conventions. This was intended to ensure that all markers followed the same procedures when scoring and recording the seven scores for each script. The second part was the review of the purpose specific rubrics. Each panel of markers familiarised themselves with the rubric for the purpose to which their panel was assigned. The descriptors for each scoring category (i.e., audience awareness and purpose, content, structure or organisation, language resources, grammar, punctuation, and spelling) were reviewed, and

discussed so that each member of the team had a similar understanding of the rubric. Three randomly selected training scripts were then marked, each followed by discussion amongst the team members to determine an appropriate score for each category, and how the rubric should be applied to that script. These training scripts and their agreed upon scores were used by panellists throughout the workshop in conjunction with the rubrics as benchmark guidelines for the implementation of the rubrics. Once each panel was confident in its common understanding, full scoring began; for most panels this was immediately after the lunch break of the first day.

Accuracy Monitoring

In addition to the initial training period, continuous monitoring of scoring agreement was implemented. Two main techniques were used: (a) check marking by the panel leader and (b) common script scoring. In the former case, each team member was asked, once an hour, to independently mark a script and record the scores on a separate record sheet for comparison with the scores assigned to the same script by the panel leader. The panel leader then marked the script, and recorded their scores on the front cover. The two sets of scores were compared for agreement, with scores within adjacent categories as the criterion. Scores outside this criterion that could not be resolved by consensus were referred to the asTTle supervisor—no such cases were reported. In the second case, a randomly selected script, appropriate to each purpose, was duplicated and issued to panellists for marking before marking began on Days Two through Four. Panellists marked the common script independently using the appropriate rubric, and a panel discussion ensued to ensure calibration of scoring for the day's marking. The independent, individual scores were collected and analysed for agreement and consistency each day.

Accuracy Estimation with the Daily Common Script

Three statistical approaches to reliability estimation of scoring are commonly undertaken; consensus, consistency, and measurement. Consensus estimates indicate the degree to which markers give the same scores, consistency estimates indicate the degree to which the pattern of high and low scores is similar among markers, and measurement estimates indicate

the degree to which scores can be attributed to common scoring rather than to error components (Stemler, 2004). As in previous asTTle marking panels (Brown, Glasswell, & Harland, in press), all three approaches were used to monitor and estimate marking reliability.

Consensus estimates. The percent adjacent agreement (i.e., what percentage of scores within plus or minus one score category of each other) consensus approach was used to estimate the degree that markers gave the same score for each script. Table 2 shows the number of markers who gave the same or within one score mark as the modal score for each of the seven score dimensions within each purpose. In this way, the percentage of adjacent agreement can be calculated for each purpose and each day.

The grand average of adjacent agreement scoring for 49 scores per each of three rating days was 69%, with excellent rates (i.e., >80%) of agreement being seen only twice, and unsatisfactory rates (i.e., <50%) of agreement seen only once for Instruct script on Day Two. The average consensus rating declined slightly over time and, although consistent with declining consensus patterns seen in large-scale marking panels (Weigle, 2002), this reduction is quite minimal.

It is assumed that the group's average score is a reasonable estimate of the script's true scores and that the size of the variance around each group's average scores for the three common scripts is a good basis for further understanding marker consensus. The curriculum sub-scores were converted to a numerical scale on the basis that Level 2 Basic represented 1, Level 2 Proficient was given 2, Level 2 Advanced equalled 3, and so on until greater than Level 6 Advanced was given 16. For each score dimension the average and standard deviation of awarded scores was calculated. The sum of a marker's difference to the average across the seven scores was found and turned into an effect size by dividing it by the average standard deviation of the seven scores.

Table 2.

Number of Markers Within One of Modal Score by Purpose by Day by Scoring Dimension

Dimension	Purpose							<i>M</i>
	Persuade	Instruct	Narrate	Describe	Explain	Recount	Analyse	
Day 1								
Markers	6	6	6	7	7	7	7	
Audience	4	5	3	5	3	4	5	
Content	5	6	5	4	4	6	5	
Structure	3	5	4	6	3	5	5	
Language	5	6	3	4	6	4	7	
Grammar	6	5	3	5	7	5	5	
Punctuation	5	5	5	4	4	5	3	
Spelling	5	5	4	4	3	5	6	
<i>M</i>	79%	88%	64%	65%	61%	69%	73%	71%
Day 2								
Markers	6	6	6	7	7	7	7	
Audience	3	3	5	5	3	4	5	
Content	4	3	4	6	4	5	5	
Structure	5	3	5	4	5	7	5	
Language	5	2	5	5	6	6	5	
Grammar	5	5	5	5	3	6	5	
Punctuation	3	2	5	4	3	5	5	
Spelling	6	3	4	6	5	7	4	
<i>M</i>	74%	50%	79%	71%	59%	82%	69%	69%
Day 3								
Markers	6	6	6	7	7	6	6	
Audience	4	4	5	6	3	3	3	
Content	3	5	6	6	3	3	6	
Structure	4	3	4	5	5	5	5	
Language	2	3	3	6	6	3	5	
Grammar	5	5	4	4	4	5	3	
Punctuation	3	5	3	5	5	4	5	
Spelling	4	5	4	5	3	5	5	
<i>M</i>	60%	71%	69%	76%	59%	67%	65%	67%
Grand Mean	71%	70%	71%	71%	60%	73%	69%	69%

An effect size of 1.0 represents on average a difference from the group average of one standard deviation of curriculum level scores. If scoring approaches the expected tolerance of within one curriculum sub-level, then the effect size should be 1.0 or less and the standard deviation should be one curriculum sub-level or less. In those circumstances, two-thirds of all scores would fall within the tolerance range of plus or minus one curriculum sub-level of the group mean. Table 3 displays the average effect size (i.e., distance from group average as proportion of average standard deviation) per marker across the three common scripts and the amount of curriculum level difference that distance represents.

Table 3.

Average Marker Difference to Group Mean in Effect Sizes and Curriculum Levels

Purpose	Marker Difference							Purpose Scores		
	1	2	3	4	5	6	7	Mean	SD	
<i>Persuade</i>										
Effect Size	0.44	0.94	0.91	0.81	0.52	0.84			7.49	1.38
Curriculum Level	0.61	1.29	1.25	1.12	0.72	1.15		1.02		
<i>Instruct</i>										
Effect Size	0.76	0.43	1.03	0.62	0.79	0.99			6.47	1.77
Curriculum Level	1.34	0.75	1.82	1.10	1.40	1.75		1.36		
<i>Narrate</i>										
Effect Size	0.54	0.70	0.55	0.59	0.90	0.58	1.05		8.67	1.86
Curriculum Level	1.00	1.29	1.03	1.09	1.67	1.07	1.95	1.30		
<i>Describe</i>										
Effect Size	0.63	0.99	0.77	0.61	0.93	0.78	0.78		10.24	1.46
Curriculum Level	0.91	1.44	1.11	0.88	1.36	1.13	1.14	1.14		
<i>Explain</i>										
Effect Size	0.74	0.92	0.87	0.66	0.90	0.78	0.45		9.23	1.78
Curriculum Level	1.32	1.64	1.55	1.17	1.59	1.39	0.80	1.35		
<i>Recount</i>										
Effect Size	1.00	0.55	0.59	0.56	0.86	0.76	0.91		9.69	1.26
Curriculum Level	1.26	0.69	0.74	0.70	1.09	0.95	1.15	.94		
<i>Analyse</i>										
Effect Size	0.57	1.01	0.81	0.64	0.87	0.89	0.36		9.90	1.49
Curriculum Level	0.85	1.51	1.21	0.95	1.29	1.33	0.54	1.10		

Marker proximity to the group mean for all seven scores across three common scripts ranged from .36 to 1.05 standard deviations, with an average of .75 *SD*. When converted to curriculum levels, two-thirds of all scores fell between .94 and 1.36 ($M = 1.12$) curriculum level sub-scores from the group average. Thus, there is a very strong central tendency for scores to fall within the adjacent-agreement consensus expectation.

Consistency estimates. Cronbach’s alpha for multiple judges was used to identify whether the pool of raters gave high and low scores in a similar pattern across the seven scores for each day’s common script. Remember that consistency estimates do not equate to raters giving the same mean scores, just that they recognise stronger and weaker scripts in a similar fashion. In order to avoid artificial deflation of correlations due to high rates of consensus, consistency estimates were calculated for all three common scripts simultaneously. In other words, each rater provided 21 matched scores which were then tested for degree of correlation (Table 4).

Table 4.

Consistency Estimates Across Three Scripts by Purpose

Purpose	Total Ratings	Consistency Estimate (alpha)
Persuade	20	.89
Instruct	18	.94
Narrate	18	.90
Describe	21	.91
Explain	20	.83
Recount	20	.97
Analyse	18	.97
<i>Average</i>	<i>19.29</i>	<i>.92</i>

The Cronbach alpha correlations show high rates of consistency among raters with, on average, over 81% of variance explained by raters giving high and low scores to the common scripts in a similar fashion. These estimates represent extremely high values for the scoring of writing (Brown, Glasswell, & Harland, in press).

Measurement estimates. Both consensus and consistency estimates are unable to properly take into account error when determining the degree of agreement in a set of scores. Measurement estimates apportion variance in assigned scores to task, rater, error, and interaction components and thus provides the most robust indicator of degree of agreement attributable to the similarity of raters' scores (Shavelson & Webb, 1991). The seven scores for three scripts scored by the six or seven raters for each purpose were analysed to generate the Brennan and Kane Dependability Index phi (Table 5). It was not possible to obtain results for day-by-day data as sample sizes were too restricted. The Brennan and Kane Dependability Index (ϕ) is calculated by obtaining the between-subjects effects error mean square, and dividing it by the sum of the absolute error variance of the set of ratings and itself. Values greater than .80 are considered dependable (Shavelson & Webb, 1991). Clearly all the scoring for the three scripts of each purpose meet the requirement of dependable measurement, with extremely consistent values seen for Persuade, Instruct, Recount, and Analyse purposes (i.e., $\phi > .95$).

Table 5.

Measurement Estimates Across Three Scripts by Purpose

Purpose	Raters	Scripts	Scores	σ^2_i	Σ^2_p	$\sigma^2_{pi.e}$	Σ^2_{ABS}	Dependability ϕ
Persuade	6	3	7	12.606	94.135	2.274	2.126	.98
Instruct	6	3	7	7.111	46.397	2.008	1.303	.97
Narrate	6	3	7	24.810	60.341	5.878	4.384	.93
Describe	7	3	7	37.150	28.190	2.421	5.653	.83
Explain	7	3	7	19.703	28.886	3.933	3.377	.90
Recount	7	3	7	4.262	56.672	1.770	0.862	.99
Analyse	7	3	7	4.928	23.163	1.736	0.952	.96
<i>Average</i>								<i>.94</i>

Users can have confidence that the sample of scripts marked in this fashion represent dependable scoring. The discrepancy between the relatively low consensus estimates and the extremely high consistency and measurement statistics suggests that the markers, while not agreeing with each other, were remarkably consistent.

Workshop Evaluation

The teachers were asked to evaluate their participation in the asTTle™ marking workshop through completion of a brief questionnaire (Table 6). Participants indicated their level of agreement with each statement using a positively packed, six-point agreement scale that has been shown to provide robust measurement (Brown, 2004). Participants were also asked to provide free response comments both orally and on the questionnaire. The three main areas of comments related to the physical conditions in which the workshop was held, the marking activities of the workshop, and participants’ anticipated benefits from having been part of the asTTle™ workshop.

Physical Environment

The physical environment for the workshop was identified by statements 4 and 9 in the questionnaire, along with additional comments. On the whole the participants were positive about the venue ($M = 5.4, SD = .7$), although there were two negative comments about the chairs. The food and beverages ($M = 5.7, SD = .6$), were largely very positively received (e.g., *“Thanks for the fantastic food – never has a catering team catered to my specific dietary needs so well. Please pass on my grateful appreciation to the chef and wait staff for being absolutely*

fantastic”). However one person requested that there be “*more fruit provided as the food was too fatty and high in cream and carbohydrates*”.

Table 6.

Ratings of Teacher Participant Satisfaction

Statement	M (SD)
1. My work was explained and directed well.	5.6 (.6)
2. This workshop gave me a clearer understanding of asTTle.	5.7 (.5)
3. Marking instructions were clear.	5.1 (.9)
4. The location was conducive to our work.	5.4 (.7)
5. Understandings gained from this workshop about writing levels will help me plan my teaching better.	5.1 (1.1)
6. The workload of the workshop was just right.	4.6 (1.3)
7. I will encourage my colleagues to participate in Project asTTle activities next year.	5.0 (1.2)
8. The rubrics used were consistent with curriculum levels 2 to 6.	4.6 (.9)
9. The food and beverages were satisfactory.	5.7 (.6)
10. This workshop has improved my understanding of what is needed to do quality assessment.	5.3 (.7)

Workshop activities

Statements 1, 2, 3, 6, and 8 pertain to the professional development impact of the workshop activities, The participants were extremely positive towards the explanation of the work (Statement 1), the clarity of instructions (Statement 3), and the clarity given them about Project asTTle™ (Statement 2). One commented “*Well organised and was really enjoyable. I was impressed by the project and leadership of the panel.*” Note that the lowest agreement amongst participants referred to the workload of the workshop (Statement 6), suggesting that participants had a small tendency to view the expectation that they should mark 6 to 8 scripts per hour as too high. The weak positive agreement that the rubrics described curriculum levels 2 to 6 (Statement 8) is of some concern. One marker commented that “*the rubrics need to be more consistently worded, with clearer steps between levels and progression more clearly delineated*”. It is expected that the revisions offered in this report, being based on user feedback, are a step in this direction.

Anticipated benefits of asTTle

This last set of statements (5, 7 and 10) asks the participants their thoughts about the use of asTTle and asTTle related activities after the completion of the workshop. Participants agreed that they would encourage their colleagues to be involved in asTTle projects in the future (Statement 7), that their experience would help them plan and assess writing better (Statements 5 and 10 respectively). One panellist provided a somewhat more extensive comment on her experience as a table leader:

I wanted to be there. It was another week of brain workout and another valuable learning experience for me. The opportunity to once again connect with other teachers from around the city and discuss the intricacies of the rubric and its application was incredibly helpful for my own teaching. Interesting also, was the chance to get a feel for student performance across the whole country. asTTle couldn't let me get away with anything too cruisy however, so this time the biggest challenge came in the form of the check marking role I was given. This involved a sensitive mix of both team work and confrontation. I was not the expert and the rubric was still not perfect, both making for interesting debates. Intensity is again the word I would use for this experience. A week of marking and checking and dialoguing with colleagues was incredibly demanding. The whole asTTle Writing experience has been one of learning and growing.

Discussion

Feedback was obtained on the rubrics and the prompts through teachers' actual use of the indicators to score student writing and on teachers' evaluation of student responses to the writing prompts. The finalised progress indicators refined by panellist feedback are displayed in Appendix 1. Changes are relatively small relative to those presented in Technical Report 45, except in the area of surface features and the addition of descriptors for Levels 2 and 3 of Analyse. It is also worth noting that these indicators have some substantial differences to those developed and used for curriculum levels 2 to 4; this is especially noticeable in the movement of level 4 indicators in asTTle V3 to levels 5 and 6 in this set. Arguments for the validity of this change were made in Technical Report 45. Note that post-marking, descriptors were generated by the Analyse panel leader and the asTTle™ team for Levels 2 and 3 of the analyse purpose. These indicators are provided to assist teachers with consistent use of the asTTle™ application and norms will be provided by linear regression based on the data available from the scoring of

a thousand scripts. Use of the analyse purpose with students who are known not to be writing at Level 4 in other purposes is not encouraged as these tasks are considered inappropriate for such students.

Results from the consensus, consistency, and measurement approaches to determining rater reliability clearly support the assertion that the teachers used the rubrics in a very similar manner; in effect, scoring as if they were one person. This means that the norms to be published in asTTle V4 can be taken as robust and that the progress indicators will have appropriate meaning to secondary school users of asTTle. Furthermore, it is evident that the high reliability of scoring is also a consequence of the procedures used to train and monitor markers. At the start of the workshop a 2 hour briefing session was provided, a further 2 hour by purpose discussion and self-training in the marking of some guinea pig scripts was conducted, hourly monitoring of each marker combined with feedback by table leaders was implemented, and daily recalibration through the marking of a common script was put into practice. These techniques for obtaining reliable scoring need to be taken seriously by any school or professional development agency taking up the challenge of meaningful assessment of writing. The training materials and procedures outlined in this document are offered as potential resource materials for such groups.

This report documents success in achieving the goals of (a) completing the marking of 7,000 scripts, (b) improving the progress indicators, and (c) refining the writing prompts. A high level of accuracy in the scoring of writing has been demonstrated. Users of asTTle™ data can have confidence in the values on which analyses and interpretations are based. Further, the report has shown that techniques used in this panel form the basis for accurate scoring in school-based applications.

References

Brown, G.T.L. (2004). Measuring attitude with positively packed self-report ratings: Comparison of agreement and frequency scales. *Psychological Reports, 94*, (1015-1024).

Brown, G.T.L., Glasswell, K., & Harland, D. (in press). Accuracy in the scoring of writing: Studies of reliability and validity using a New Zealand writing assessment system. *Assessing Writing*.

Brown, G.T.L., Pipe, D., & Higginson, R. (2003, December). *Writing at Level 5—6 of the New Zealand English Curriculum: Development of Progress Indicators & Tasks*. (asTTle Tech. Rep. 45). University of Auckland/Ministry of Education.

Glasswell, K., & Brown, G. T. L. (2003, August). *Accuracy in the scoring of writing: Study in large-scale scoring of asTTle writing assessments*. (asTTle Tech. Rep. 26). University of Auckland/Ministry of Education.

Glasswell, K., Parr, J., & Aikman, M. (2001). *Development of the asTTle writing assessment rubrics for scoring extended writing tasks*. (asTTle Tech. Rep. No. 6). Auckland, NZ: University of Auckland/Ministry of Education.

Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research & Evaluation, 9*(4).

Shavelson, R. J., & Webb, N. M. (1991). *Generalizability Theory. A Primer*. Newbury Park, CA: Sage.

Weigle, S.C. (2002). *Assessing writing*. Cambridge, UK: Cambridge University Press.

Appendix 1. Progress Indicators Levels 2—6

Persuade (<i>to argue for or influence opinion</i>)					
Dimension	Level 2 Proficient	Level 3 Proficient	Level 4 Proficient	Level 5 Proficient	Level 6 Proficient
Audience Awareness and Purpose	<ul style="list-style-type: none"> Evidence that writer recognises that his/her opinion is needed. May state opinions from a personal perspective. 	<ul style="list-style-type: none"> Language generally appropriate to audience. Some attempt to influence the reader is evident. 	<ul style="list-style-type: none"> Writer aware audience may hold a different point of view but tends to assume there is only one different generalised point of view Opening presents a point of view to audience Writing attempts to persuade reader Clearly stated consistent position is evident. 	<ul style="list-style-type: none"> Identifies & relates to a concrete / specific audience Awareness of intended reader evident at beginning & end, but may be inconsistent in middle sections Language use is appropriate and has elements which begin to be persuasive to audience 	<ul style="list-style-type: none"> Implicit awareness that audience may hold a range of points of view Consistently persuasive throughout for intended audience Tone likely to impact on or affect change or manipulate the intended audience towards author's purpose
Content / Ideas	<ul style="list-style-type: none"> Writing covers some (1 or more) task and topic appropriate domains: (e.g. position statement – writer identifies position on the issue, makes 2 or more simple opinion/statements related to the topic, makes use of a final statement to round off the text in some way). Can include many statements unrelated to the topic and/or task. 	<ul style="list-style-type: none"> Most argument domain elements are present (i.e., main points, some supporting evidence, or illustration, re-statement of position). Some elaboration of main points occurs. May include information that does not contribute to argument. 	<ul style="list-style-type: none"> Argument domain elements (e.g. position statement, main points, illustration / evidence of main points, restatement) are identifiably present, yet simple. List-like presentation of ideas. Some relevant support for ideas may be evident but is inadequate or inconsistent. Conclusion restates position. 	<ul style="list-style-type: none"> Argument domain elements (e.g. plausible position statement, main points, illustration, evidence of main points, restatement) are somewhat developed, usually consistent, and nearing comprehensive. Support / reasons show strong link to argument. Conclusion reflects points made and may expand the argument 	<ul style="list-style-type: none"> Examples are well chosen to support writer's purpose. Consistent support given to main points. Content and points made are considered, selected, relevant and elaborated Content adds weight to the writer's position. Conclusion provides thematic integration of argument, rather than simply repeating or summarising points made

Persuade *(to argue for or influence opinion)*

Dimension	Level 2 Proficient	Level 3 Proficient	Level 4 Proficient	Level 5 Proficient	Level 6 Proficient
Structure / Organisation	<ul style="list-style-type: none"> • Semblance of organisation is evident (e.g., some grouping of ideas) • Text may be limited because of presentation of opinion statements as discrete elements. 	<ul style="list-style-type: none"> • Evidence of attempts at overall structuring of content through grouping ideas within and across sentences • May use devices such as paragraphing and simple linking of ideas through conjunctions such as because, and, since, although etc. 	<ul style="list-style-type: none"> • Content is grouped logically at main idea level • Ideas not necessarily sequenced persuasively • Linkages simple if present 	<ul style="list-style-type: none"> • Content managed through grouping and/or paragraphing main ideas & supporting evidence. • Overall structure evident but may not contribute to intended effect or argument. • Paragraphs structured with main ideas and support for argument. • Simple connectives /linkages present but inconsistently used • Relevant material is selected to support identified key ideas (and to persuade the audience) 	<ul style="list-style-type: none"> • Explicit, logical structure enhances argument. • Logical arrangement of reasoned ideas in well crafted paragraphs and the use of strong topic sentences to guide the reader's understanding of the writer's argument • Consistent use of a variety of connectives / linkages (e.g., on the one hand, however, although) • Evidence of more complex linkages (e.g. varied use of linking words and phrases, conjunctions and text connectives).

Persuade (*to argue for or influence opinion*)

Dimension	Level 2 Proficient	Level 3 Proficient	Level 4 Proficient	Level 5 Proficient	Level 6 Proficient
Language Resources / Choices	<ul style="list-style-type: none"> • Language has structure of simple opinion statements, but may be stated from a personal perspective (e.g., "I reckon"). • Topic or content related language present but little opinion is conveyed through language choices (e.g., nouns may be neutral, may have basic descriptors, or may lack simple adjectivals. Verbs and adverbials may be limited). • Shows some understanding of the use of pronouns but reference may be unclear or overused. • Simple sentences used but may attempt complex sentences 	<ul style="list-style-type: none"> • Evidence of some task appropriate structures and language (e.g., attempts to use verbs in passive structures to make arguments seem more objective and convincing). • Evidence that the writer is a beginning to select language to create a particular effect and to influence the reader (e.g. "point of view " nouns, viewpoint adverbials, opinion adjectives, adverbs and adjectives to add detail and weight to opinion statements and evidence, some use of modal auxiliary verbs (can, might, should, may) present). • May be some unclear or repetitious reference. • Many simple sentences; may have some complex sentences. 	<ul style="list-style-type: none"> • Language is appropriate, yet straight-forward • Language identifies a particular viewpoint and is used to persuade the reader (e.g. "point of view" nouns, viewpoint adverbials, opinion adjectives, adverbs, & adjectives to add detail and weight to opinion statements and supporting evidence). • Pronoun reference links clear. • Uses complete sentences. 	<ul style="list-style-type: none"> • Use of persuasive language techniques for the deliberate effect of involving intended reader in and persuading them of the argument being put • Consistent use of appropriate language for task and topic • Appropriate use of passive structures and modal auxiliaries to strengthen argument • Complex sentences evident, where appropriate. 	<ul style="list-style-type: none"> • Appropriate use of vocabulary to task, purpose, and audience • Language choices are considered and show selection for effect with intention to manipulate / influence audience • Tone (e.g., sarcastic, threatening, humorous, emotive, etc.) underpins language selection and strengthens argument

Instruct (*to instruct or layout a procedure*)

Dimension	Level 2 Proficient	Level 3 Proficient	Level 4 Proficient	Level 5 Proficient	Level 6 Proficient
Audience Awareness and Purpose	<ul style="list-style-type: none"> Evidence that the writer recognises the purpose for writing (e.g., attempts to instruct the reader) and that writing is for an audience other than the self (e.g., uses “you” or imperative or command form). May assume shared knowledge with the reader. 	<ul style="list-style-type: none"> Language use and writing style is appropriate to audience. Instructs but relies on context Requires some reader inference to complete the procedure. 	<ul style="list-style-type: none"> Little audience inference is required for simple task to be done. Writer may not make sufficient allowance for prior knowledge of intended audience. 	<ul style="list-style-type: none"> Personal experience of writer may form basis of rationale for instructions. Some awareness of audience having different degrees of prior experience with task (e.g., may have already done this task or similar or never done such a task before). Speaks directly to audience. Little audience inference required for complex instructions. 	<ul style="list-style-type: none"> Experience, background, purpose, motivation, needs (including psychological and emotional) of audience taken into account. Gives / explains audience-appropriate rationale for instruction. Empathy shown for intended reader’s needs relative to completing instructions in real world.
Content / Ideas	<ul style="list-style-type: none"> Some elements of procedure (e.g., headings, actions, materials) included. Some topic-related information included. Evidence of instruction-like statements. 	<ul style="list-style-type: none"> All basic procedure elements (i.e., headings, actions, materials) included. Some elaboration of elements. Limited off-topic information. 	<ul style="list-style-type: none"> Most content is relevant. Simple task can be completed because enough content / accuracy / precision are provided. List-like instructions. 	<ul style="list-style-type: none"> Content wholly relevant. Provides general advice on method/ procedure/ support & rationale. Procedure elements sufficiently elaborate, precise, and comprehensive (e.g., sub-steps elaborated, diagram / illustration as appropriate). Procedure elements (i.e., headings, sub-headings, materials, actions) show developed content knowledge. Main and supporting ideas given with evidence of selection where appropriate. 	<ul style="list-style-type: none"> Targeted advice provided for foreseen difficulties. Running commentary or rationale provided to give psychological and emotional support. Effective and judicious excision or selection of detail (not over-detailed). Conditional steps included.

Instruct (*to instruct or layout a procedure*)

Dimension	Level 2 Proficient	Level 3 Proficient	Level 4 Proficient	Level 5 Proficient	Level 6 Proficient
Structure / Organisation	<ul style="list-style-type: none"> • Semblance of order to procedure. • May use a simple ordering device (e.g., numbers). 	<ul style="list-style-type: none"> • Some grouping or sequencing of procedure elements evident. • Some use of ordering devices but limited in number and/or scope (e.g., repetition of then). 	<ul style="list-style-type: none"> • Adequate grouping of content. • Adequate use of headings and numbering. • Conventional paragraphing. • Elements of procedure grouped or sequenced appropriately. 	<ul style="list-style-type: none"> • Clear, logical, coherent structure • Content tailored to audience and task needs rather than providing all possible relevant material on topic. • Deliberate use of ordering devices and combinations of organisational method may be used. Sequence indicated in paragraph links. • Uses (and derives from task) categories, headings, and subheadings effectively. • Introduction sets tone, gives parameters, and justification for writing. • Expansions placed logically. 	<ul style="list-style-type: none"> • Clear, logical, coherent structure but may extend / manipulate conventional structures. • Strong links between paragraphs used to hold readers attention. • Variety of effective, appropriate methods used to organise material (e.g., itinerary, list, subject group paragraphs). • Effective concluding address to reader.
Language Resources / Choices	<ul style="list-style-type: none"> • Simple, unelaborated statements evident. • Some command-like statements present. • Actions recounted from a personal perspective. • Simple sentences used but may attempt complex sentences 	<ul style="list-style-type: none"> • Evidence of use of task appropriate language (e.g., use of descriptors – action verbs, adverbs, adjectives) to describe materials and actions. • Command-like statements predominate. • Attempts to use generalised other (e.g., second or third person). • Many simple sentences correct. • Tendency to over-use compound 'run-on' sentences • Some complex sentences evident 	<ul style="list-style-type: none"> • Consistently refers to reader in generalised way (one/you). • Language use and writing style directed to audience. • Some use of task-appropriate language (use of descriptors – action verbs, adverbs, adjectives) to clarify procedure. • Complex sentences evident, where appropriate. • Uses complete sentences. 	<ul style="list-style-type: none"> • Appropriate use of specialised language evident but not always sustained. • Topic specific language used. • Shows a variety of address which develops relationship with reader depending on purpose of text. • May adjust language to both instruct and advise. • Evidence of concise use of language. • Conclusion may use an imperative. 	<ul style="list-style-type: none"> • Controlled and consistent use of appropriate register. • Rich in adjectives / positives to motivate and support completion of instructions. • Complex, appropriate, varied sentence selection.

Narrate (*to inform or entertain through imaginative narrative*)

Dimension	Level 2 Proficient	Level 3 Proficient	Level 4 Proficient	Level 5 Proficient	Level 6 Proficient
Audience Awareness and Purpose	<ul style="list-style-type: none"> Evidence that the writer recognises the purpose for writing (to tell a story) and that he/she is writing for an audience other than the self. 	<ul style="list-style-type: none"> Attempts to capture the intended reader's interest. Language use and writing style appropriate to telling a story. Attempts to adopt a perspective to tell the story. 	<ul style="list-style-type: none"> The text attempts to entertain Beginning of text attracts attention and provides adequate context for narrative Attempts internally consistent, credible world for reader to enter 	<ul style="list-style-type: none"> Audience is entertained by text but not consistently. Mostly credible and internally consistent world 	<ul style="list-style-type: none"> Text entertains, engages, challenges, provokes, etc. audience and sustains reader attention. Sustained, credible world; largely suspends reader's disbelief
Content / Ideas	<ul style="list-style-type: none"> Some attempt at a story. Writing is a series of loosely related sentences or a series of sentences that all describe a single event. 	<ul style="list-style-type: none"> Writing includes some important elements of story (e.g., has essentials of character, settings, and events). Evidence of inclusion of problem or complication. May attempt to conclude events. 	<ul style="list-style-type: none"> Text contains all important elements (e.g., orientation, complication, resolution, and sometimes coda) but often the ending is inadequately controlled. Story is largely plot driven Characters and setting are developed to a skeletal level only. Content may be undisciplined; lacks selection, organisation and interest for reader (dumping) 	<ul style="list-style-type: none"> Where traditional narrative used, necessary elements are present Some focus on and development of specific events, characters, or settings. Ending may reflect or tie in with orientation 	<ul style="list-style-type: none"> Clear focus on and appropriate development of specific events, characters, and settings. Ending provides completion for both traditional and non-traditional structures Clear thematic connection between beginning and ending
Structure / Organisation	<ul style="list-style-type: none"> Semblance of order evident but limited because of haphazard or stream of consciousness-type organisation. 	<ul style="list-style-type: none"> Some arranging of story elements evident. The story is organised around happenings and has a point. Ideas/ events may be linked through the use of devices such as paragraphing or linking words and/or phrases (e.g., later that evening, because). 	<ul style="list-style-type: none"> Simple story element arrangement managed well, yet may have a sense of disjunction between beginning and end. Use of simple linking devices (e.g., conjunctions of time: (after, next, later, then, etc.) make the story progress. 	<ul style="list-style-type: none"> Complex story element arrangement managed inconsistently (e.g., effective plot or development of events, but character development incomplete). Inconsistent control of pace and proportion of elements. May experiment with non-traditional structures (e.g., slice-of-life) with limited success 	<ul style="list-style-type: none"> Complex story element arrangement largely sustained (e.g., effective plot or development of events without sacrificing character development). May attempt non-traditional story structures with increasing success Pacing and proportion largely controlled. Effective linking is evident

Narrate (*to inform or entertain through imaginative narrative*)

Dimension	Level 2 Proficient	Level 3 Proficient	Level 4 Proficient	Level 5 Proficient	Level 6 Proficient
Language Resources / Choices	<ul style="list-style-type: none"> • Language is simple. • Actions recounted with little elaboration and, overall, style lacks variety (may be limited for topic e.g. pedestrian use of descriptors – adverbials, adjectives – like nice/ly). • May insert direct speech but context lacks clarity. • Simple and compound sentences used; may attempt complex sentences. • May have little variety in sentence types 	<ul style="list-style-type: none"> • Evidence of attempts to add interest and detail through the use of descriptors (e.g., adverbials, adjectives). • May attempt to use dialogue to add to story. • Many simple sentences correct. • Some complex sentences evident • Attempts to vary sentence types 	<ul style="list-style-type: none"> • May use language devices (e.g., figurative language) and descriptors (e.g., adverbials, adjectives) to give detail to plot • Uncontrolled use of dialogue (where included). • Some variation of sentence types. • Complex sentences evident, but may be inappropriately long. • Some attempt to experiment with vocabulary, though largely restricted to a relatively limited range (up to 5,000 most common words) 	<ul style="list-style-type: none"> • May use language and descriptors (e.g., adverbials, adjectives) to give detail to plot, characters, and settings • Purposeful use of dialogue (where included). • Some effective variation of sentence types. • Complex sentences evident and of appropriate length. • Some appropriate matching of vocabulary to narrative • The writer’s style is evident in parts of the text. 	<ul style="list-style-type: none"> • Largely controlled use of descriptive or figurative language to show physical, emotional, psychological states rather than just naming them (“show not tell”) • Largely appropriate, effective, varied range of sentence choices. • Vocabulary and language consistently appropriate to narrative • The writer’s style permeates the text and may address reader directly • Largely sustained control of a range of story elements including dialogue, description, tension, emotion, pace, etc.

Describe (to classify, organise, describe, and report information)

Dimension	Level 2 Proficient	Level 3 Proficient	Level 4 Proficient	Level 5 Proficient	Level 6 Proficient
Audience Awareness and Purpose	<ul style="list-style-type: none"> Evidence that the writer recognises the purpose for writing. Gives information from a writer's perspective (e.g., may require reader to infer or select information to make sense of complete text). 	<ul style="list-style-type: none"> Language use and writing style generally appropriate to audience and purpose. Informs but may require reader inference. 	<ul style="list-style-type: none"> Inadequate background /context provided for intended audience, except in basic or simplistic tasks. Awareness of audience/purpose most evident in introduction 	<ul style="list-style-type: none"> Provides adequate background and little reader inference is required. Audience included directly or indirectly in text and referred to at beginning and end. Text content sufficient to meet audience's needs 	<ul style="list-style-type: none"> Appropriate background and context to meet intended audience's needs Consistent / sustained reference to audience needs and own purpose throughout text.
Content / Ideas	<ul style="list-style-type: none"> Evidence of statements of fact, including facts relevant to the topic and task, covering, for example, some (2 or more) task-appropriate domains: attributes, behaviours, properties, functions, location, etc. Can include many statements unrelated to the topic and/or task. 	<ul style="list-style-type: none"> Most domain elements appropriate to the task present (e.g., the writer classifies and deals with attributes, behaviours, properties, functions, location, etc.). May include some material unrelated to the topic of the given task. 	<ul style="list-style-type: none"> Simple content and tasks handled adequately; struggles with complex content. Almost all material is related to the topic of the given task Content described is largely one-faceted Domain elements may be present for the given task but incomplete, or inconsistent 	<ul style="list-style-type: none"> Description / report answers the task set in the prompt. Domain elements are present for the given task (e.g., title, classification of content to be described or reported, uses factual statements to deal with attributes, behaviours, properties, functions, location, etc. as appropriate, makes use of a final statement to round off the text in some way). Some elaboration in description Some complexity of content or perspectives (two or more) Selects material relevant to task and audience 	<ul style="list-style-type: none"> Describes and adds some interpretation or evaluation material May include poetic detail Comprehensive, information laden, and consistent elaboration (i.e., the writer classifies, deals with attributes, behaviours, properties, functions, location, etc.)

Describe (*to classify, organise, describe, and report information*)

Dimension	Level 2 Proficient	Level 3 Proficient	Level 4 Proficient	Level 5 Proficient	Level 6 Proficient
Structure / Organisation	<ul style="list-style-type: none"> • Semblance of framework (e.g., some grouping of information). • Text may be limited because fact statements presented as discrete elements. • May be attempting to section or paragraph. 	<ul style="list-style-type: none"> • Evidence that the writer is using a framework for ordering content (e.g., categorising or classifying). • May not be consistently or optimally ordered, and elements may be inappropriately assigned to parts of framework. • Evidence of attempts at sectioning or paragraphing. 	<ul style="list-style-type: none"> • Straight-forward, simplistic, obvious frameworks for ordering report or description handled adequately. • Grouping is largely list-like • Complex, thematic structures may be attempted, but unsuccessfully handled 	<ul style="list-style-type: none"> • Clearly organised, thematic framework utilised but may be inconsistent • Introduction and conclusion are used to develop focus on topic • Paragraphs used with main ideas and supporting details. Thematic linking of sentences to topic of paragraph or section. • Elements of description appropriately assigned. 	<ul style="list-style-type: none"> • Coherent, logical, thematic structure throughout. • Introduction and conclusion enhance writer's purpose. • Description has shape - begins with general statement then moves to particular • Structure enhances and strengthens ideas in text

Describe (*to classify, organise, describe, and report information*)

Dimension	Level 2 Proficient	Level 3 Proficient	Level 4 Proficient	Level 5 Proficient	Level 6 Proficient
Language Resources / Choices	<ul style="list-style-type: none"> • Language has structure of simple factual descriptions (e.g., use of relating verbs to tell what "it is" or what "it has"). • Topic-related language present but little detail conveyed through language (e.g., nouns may have basic descriptors or may lack simple adjectivals). • Verbs may be limited and lack simple adverbials). • Shows some understanding of the use of pronouns but reference (the who or what) may be unclear or overused • Simple sentences used, but may attempt complex sentences. 	<ul style="list-style-type: none"> • Evidence of use of task appropriate structures and language (e.g., relating verbs (to be/to have) for classifying; action verbs for describing behaviours or uses (most often present tense); use of adverbs and adjectives shows an attempt to provide the reader with detail). • May be some unclear or repetitious reference. • Many simple sentences, with some complex sentences evident. 	<ul style="list-style-type: none"> • Straight-forward tasks handled with adequate command of task appropriate language • Reference links clear. Language of comparison is evident. • Use of appropriate language for task and topic evident (e.g., adding information to the noun or noun "packing"). • Evidence of attempts to add information by noun- group "packing" or by the use of complex adverbials). • Uses complete sentences, and complex sentences evident, where appropriate. 	<ul style="list-style-type: none"> • Consistent use of appropriate language for task and topic (e.g., relating verbs (to be /to have) for classifying; action verbs for describing behaviours or uses (most often use present tense) • Competent plain descriptive prose used successfully • Some figurative language used, but may be inappropriate or uncontrolled "purple prose" • Attempts to give vivid description and use technical and less-frequent vocabulary • Complex sentences used appropriately 	<ul style="list-style-type: none"> • Controlled/rich vocabulary evident. • Uses a range of language techniques (incl. figurative language) deliberately to create an effect in text • Largely appropriate, varied, and purposeful sentence structures used.

Explain (to give an account with reasons of how or why something is formed or works)

Dimension	Level 2 Proficient	Level 3 Proficient	Level 4 Proficient	Level 5 Proficient	Level 6 Proficient
Audience Awareness and Purpose	<ul style="list-style-type: none"> The writer recognises that an explanation is required and that he/she is writing for an audience other than the self. Assumed shared knowledge with the reader may interfere with meaning. 	<ul style="list-style-type: none"> Language and writing style is that of explanation. It is appropriate to the audience. May rely on context and requires some reader inference to understand explanation. 	<ul style="list-style-type: none"> Explanation of a straightforward task is clear and can stand alone Language use and writing style is appropriate to the explanation and directed to the reader / audience (e.g., evidence that the reader's needs are being considered). 	<ul style="list-style-type: none"> Explanation is directed to the audience and rationale appropriate to the intended audience for explanation is given but not always developed. Task complexity may prevent writer from meeting intended audience needs. 	<ul style="list-style-type: none"> Explanation is directed to the audience and rationale appropriate to the intended audience for explanation is developed. Acknowledgment of audience is sustained throughout text and in conclusion. Intended audience needs met consistently despite task complexity.
Content / Ideas	<ul style="list-style-type: none"> Writer makes an attempt to identify the phenomenon or process and gives two or more simple reasons for its occurrence. May include statements that are unrelated to the topic and/or task or include a personal perspective to the explanation (e.g., I like rocks. I saw a tuatara at the zoo in Auckland). 	<ul style="list-style-type: none"> Writer identifies the phenomenon or process clearly. Body of text contains further elaboration and gives associated reasons for why/how it occurs. Limited unrelated information evident. 	<ul style="list-style-type: none"> Writer identifies the phenomenon process clearly, and may also give contextualising information. Body of text contains a sequenced account of straightforward aspects / processes, and gives associated reasons for 'how it occurs'. These tasks tend to be ordered or time-oriented explanations Little explanation of 'why' is proffered Includes only relevant content. 	<ul style="list-style-type: none"> Content is clear, adequately detailed, and relevant to topic sentences / paragraphs. Simplistic single chain-like immediate causes are used to explain why phenomenon occurs Generalised level of relevant, accurate details provided for each stage. Complexity of task partially handled. 	<ul style="list-style-type: none"> Content is clear, comprehensive, and relevant to topic Multiple causes / aspects of phenomenon are acknowledged and explained. Specific relevant, accurate details selected and targeted to support explanation. Complexity of task consistently handled.

Explain *(to give an account with reasons of how or why something is formed or works)*

Dimension	Level 2 Proficient	Level 3 Proficient	Level 4 Proficient	Level 5 Proficient	Level 6 Proficient
Structure / Organisation	<ul style="list-style-type: none"> • Generally organised at sentence level. • May attempt to show cause-and effect relationships in the explanation by using within-sentence links (e.g., because, so). • May be attempting to paragraph. 	<ul style="list-style-type: none"> • Evidence of attempts at structuring content through the grouping of ideas within and across sentences. • May be attempting to construct between-paragraph links. 	<ul style="list-style-type: none"> • Beginning provides orientation and may provide context. • Attempts at grouping or sequencing explanation (e.g., thematic grouping) across the text. 	<ul style="list-style-type: none"> • Straightforward conventional structure apparent (intro, body, conclusion) • Expresses causal relationships through links. • May be inconsistent sequencing within or between paragraphs 	<ul style="list-style-type: none"> • Concise introduction to topic and structured overview of student's own text given • Clear sequential structures / transitions evident within and between paragraphs - • Consistent selection, placement and ordering of relevant material in paragraphs and throughout the text
Language Resources / Choices	<ul style="list-style-type: none"> • Simple factual or declarative statements evident. • Action verb choices reflect topic and task. • Topic-related vocabulary present but little detail conveyed through the language (e.g., nouns may lack adjectivals). • Shows some understanding of pronoun use but unclear or overused reference may interfere with meaning. • Simple sentences used but may attempt complex sentences. 	<ul style="list-style-type: none"> • Evidence of task appropriate language (e.g., action verbs mainly present tense to tell how it is or happens). • Topic-related vocabulary contributes to understanding of parts of phenomenon to be explained. • Some use of adjectivals / adverbials to give detail and precision to explanation. • May be some unclear or repetitious reference. • Many simple sentences used and some complex sentences evident. 	<ul style="list-style-type: none"> • Simple linking language used (e.g., conjunctions) to connect ideas within and across sentences. Pronoun reference links clear • Conjunctions used to link ideas within and across sentences. • Technical / specialised language (jargon) is attempted • Uses complete sentences. Complex sentences evident, where appropriate. 	<ul style="list-style-type: none"> • Straightforward linking language used but may not be used consistently. • Technical / specialised language (jargon) is accurately used. • Complex sentences evident and used appropriately. • Appropriate formality attempted. 	<ul style="list-style-type: none"> • Appropriate and varied linking language sustained in use. • Variety of appropriate, purposeful sentence structures is evident (including use of passive voice). • Appropriate formality sustained.

Recount (to inform or entertain a reader by reconstructing a view of the writer's own world)

Dimension	Level 2 Proficient	Level 3 Proficient	Level 4 Proficient	Level 5 Proficient	Level 6 Proficient
Audience Awareness and Purpose	<ul style="list-style-type: none"> The writer recounts to tell of a past experience or happening. Recognises that he/she is writing for an audience other than the self, but may be limited by assumption of shared knowledge. 	<ul style="list-style-type: none"> Language use and writing style appropriate to recounting a past happening. Audience has most information needed to make sense of the experience recounted. Some inference may be needed. Evidence of attempts to capture the audience's interest. 	<ul style="list-style-type: none"> The text attempts to entertain / inform. Beginning of text attracts attention and provides adequate context for recount. The text provides sufficient description of setting, situation, etc. for audience's understanding. First person recount speaks to audience. 	<ul style="list-style-type: none"> Writer explains events, situations, etc. as appropriate for the audience May experiment with point of view or voice (e.g., use a 'hook/ quote to engage reader's interest). Writer's developing awareness of audience is seen in reflective comments on the action. 	<ul style="list-style-type: none"> Writer's awareness of audience is seen in ending of text Speaks directly / indirectly to audience making strong connections and drawing audience into recount. Adjusts content, details, and style to effectively bridge gap between writer and audience.
Content / Ideas	<ul style="list-style-type: none"> Writer recounts events. Writing begins with an orientation (background information) using some of the elements of recount (when, where, who, what, and why). May be some evidence of selection of events for inclusion or of comment on events. May include content not relevant. 	<ul style="list-style-type: none"> Writing includes evidence of foregrounding of significant content. Evidence of attempts to add detail to, comment on, or evaluate selected points of interest. There may be an attempt to conclude. 	<ul style="list-style-type: none"> Necessary elements of recount are present. Focus on and development of some specific events and interest areas. Content selected for narrative rather than thematic impact. Simple appropriate conclusion. 	<ul style="list-style-type: none"> Orientation is comprehensive, yet succinct. Clear focus on and development of specific events and interest areas but not always edited for relevance Thematic understanding / or interpretation of recount is attempted Some evidence of interpretive reflection on events in recount, possibly through sharing thoughts and feelings with the audience. 	<ul style="list-style-type: none"> Writer's sense of looking back and analysing event is integrated fully through the recount. Recount is enriched with interpretive comments, evaluation and observation, and sustained appropriate selection of details. Conclusion is linked thematically to recount content.

Recount *(to inform or entertain a reader by reconstructing a view of the writer's own world)*

Dimension	Level 2 Proficient	Level 3 Proficient	Level 4 Proficient	Level 5 Proficient	Level 6 Proficient
Structure / Organisation	<ul style="list-style-type: none"> Events are largely sequenced in time order. Events are linked by using common words that indicate the passage of time (first, then, next etc). 	<ul style="list-style-type: none"> Events are in time order and seem to follow on. Events are linked in a variety of ways (e.g., by using devices that indicate passage of time such as later, following, after that). Attempts at paragraphing into sections (e.g., orientation, sequence of events, reorientation). 	<ul style="list-style-type: none"> Events are in time order, and sequencing is managed well. Paragraphs support the structure 	<ul style="list-style-type: none"> Narrative structure is carefully used to convey and integrate writer's emotional response to events 	<ul style="list-style-type: none"> Detailed organisation of events interwoven with interpretive reflection / observation
Language Resources / Choices	<ul style="list-style-type: none"> Language is simple. Some action verbs used but limited in scope (e.g., "I went", "I got" etc.). Events and actions recounted with little elaboration (may be limited by repetitive use of sentence structure and/or language to indicate passage of time). Uses simple past tense. May insert direct speech. Simple sentences used, but may attempt complex sentences 	<ul style="list-style-type: none"> Evidence of attempts to add detail to content through using a variety of verbs, adverbials, and adjectivals. Pronoun use appropriate and consistent. Uses appropriate verb tense for content inclusions. Variety in sentence structure. May include dialogue to assist reconstruction of events. Many simple sentences--some complex sentences may be evident. 	<ul style="list-style-type: none"> May use language devices (e.g., figurative language, metaphors) and descriptor (e.g., adverbials, adjectives) to amplify content. Some controlled use of dialogue (where included). Some attempt to experiment with vocabulary, though largely restricted to a relatively limited range (up to 5,000 most common words) Uses complete sentences. Complex sentences evident. Varied use of verbs to describe actions and events and to capture thoughts and feelings. 	<ul style="list-style-type: none"> May use language devices (e.g., figurative language, metaphors) and descriptors (e.g., adverbials, adjectives) to give detail to events, persons, and experiences Some appropriate matching of vocabulary to increasing maturity of life experiences being recounted. Purposeful, appropriate use of complex sentences. 	<ul style="list-style-type: none"> Consistent appropriate matching of vocabulary to maturity of life experiences being recounted. Controlled use of descriptive / figurative language to show physical, emotional and psychological states rather than just naming them (show not tell) Largely sustained use of appropriate and varied sentences.

Analyse (to expose a critical analysis of characteristics and merits of texts)

Dimension	Level 2 Proficient	Level 3 Proficient	Level 4 Proficient	Level 5 Proficient	Level 6 Proficient
Audience Awareness and Purpose	<ul style="list-style-type: none"> Text written for self rather than for an audience, although may have token recognition of audience (e.g., occasional 'you') Tendency to express purpose as description rather than analysis 	<ul style="list-style-type: none"> Awareness of audience has some impact on language or writing style, including direct address to audience (i.e., 'you') Largely descriptive intention, with some elements of analysis identified 	<ul style="list-style-type: none"> Language use and writing style is appropriate to the audience. Identifies relevant key elements of text being analysed. 	<ul style="list-style-type: none"> Audience's expertise is acknowledged but this may not always be sustained. Explains relevance of key elements of text to task 	<ul style="list-style-type: none"> Audience's expertise is acknowledged by largely sustained, appropriate selection and interpretation of content. Evaluates merit/worth of relevant key elements
Content / Ideas	<ul style="list-style-type: none"> Text is largely a simplified repetition of plot elements only; with strong emphasis on personal experience and reaction Much irrelevant or repetitive material included Generally too brief for adequate description of relevant text elements 	<ul style="list-style-type: none"> Largely a description of text elements; may have some personal response Description of text generally skeletal and inadequate, including repetition and irrelevance. May identify 1-2 point(s) suitable for analysis Generally too brief to provide basis for analysis 	<ul style="list-style-type: none"> Personal response rather than analysis predominates. Straightforward description of text is given. Student text may include unnecessary repetition or retelling of text. Some relevant text analysis attempted 	<ul style="list-style-type: none"> Uses text analysis (e.g., plot, character, theme, setting) to explain topic. Relevant details (incl. quotations) are given and writer attempts to explain their importance to their analysis. Some use of appropriate personal response or reference to other texts. 	<ul style="list-style-type: none"> Uses text analysis (i.e., character, theme, setting, language, or content) to interpret topic. Use of appropriate personal responses, analogies, and / or reference to other texts. Integration of textual evidence into writing. Details (incl. quotations) are used consistently and appropriately.
Structure / Organisation	<ul style="list-style-type: none"> Organisation largely jumbled with little deliberate sequencing of the content Structure may reflect naïve stream of consciousness related to personal response or reflect order of events within text being analysed 	<ul style="list-style-type: none"> Organisation largely driven by plot sequence of source text, though tendency for unsupported jumps. Ideas are sequenced logically in paragraphs, though probably not displayed conventionally. 	<ul style="list-style-type: none"> Beginning provides orientation and may provide context for writing Attempts are evident of grouping or sequencing of analysis (e.g., thematic grouping). Across the text there is a sense of an attempt to organise content. 	<ul style="list-style-type: none"> Straightforward conventional structure is apparent (introduction, body, and conclusion). Ideas are structured and organised in paragraphs with topic sentences to signpost key ideas. Some appropriate ordering and linking between paragraphs is evident. 	<ul style="list-style-type: none"> Concise introduction to topic and structure overview of student's own text is given. Strong organisational structures / transitions are evident within and between paragraphs. Consistent selection, placement, and ordering of material in paragraphs and throughout the text.

Analyse (*to expose a critical analysis of characteristics and merits of texts*)



Dimension	Level 2 Proficient	Level 3 Proficient	Level 4 Proficient	Level 5 Proficient	Level 6 Proficient
Language Resources / Choices	<ul style="list-style-type: none"> • Emotive language used descriptively • Vocabulary choices restricted to simple statements without elaboration or modification • Most transitions not signalled or introduced. 	<ul style="list-style-type: none"> • Inappropriate subjective and even emotive language predominates. • Random and inappropriate use of specialised literary terminology but generally appropriate vocabulary is selected. • Some intermittent use of linking language • Simple sentences and run-on structures dominate 	<ul style="list-style-type: none"> • Writer uses appropriate language to give clarity to the analysis. • Attempts to use specialised literary terminology appropriately. • Simple linking language is used (e.g., conjunctions) to link ideas within and across sentences. • May be restricted to simple sentences and repetitive structures. 	<ul style="list-style-type: none"> • Appropriate formality is evident but may not be used consistently. • Basic specialised literary terminology is accurately used. • Straightforward linking language is evident but may not be used consistently. • Use of complex sentences is evident. 	<ul style="list-style-type: none"> • Appropriate formality is sustained. • Accurate use of range of specialised literary terminology. • Appropriate and varied linking language is sustained throughout text. • Largely sustained use of appropriate and varied sentences.

Surface Features *(to control the accuracy of the mechanical conventions of written language)*

Dimension	Level 2 Proficient	Level 3 Proficient	Level 4 Proficient	Level 5 Proficient	Level 6 Proficient
Grammar	<ul style="list-style-type: none"> Errors interfere with meaning (e.g., uses participle phrase; may have errors in agreement, tense, and/or word order). 	<ul style="list-style-type: none"> Some errors in sentences (e.g., subject and finite verb not present or do not agree, split infinitive, word order). Tenses may vary unintentionally but meaning is clear. 	<ul style="list-style-type: none"> Uses complete sentences Uses basic sentence types correctly (Simple, Affirmative, Active, Declarative, Negative, Interrogative) Subject-verb agreement and appropriate word order controlled for straightforward sentences Some control of compound and complex sentences Inconsistent control of tense Errors no longer interfere with meaning 	<ul style="list-style-type: none"> Control of complex sentences evident Subject-verb agreement and appropriate word order controlled for compound/complex sentences Control may not be evident in all parts of text Largely maintains consistent tense or point of view Few errors to interfere with communication 	<ul style="list-style-type: none"> Sustained control of sentence grammar evident throughout piece with only few errors Control enhances communication
Punctuation	<ul style="list-style-type: none"> Basic correct sentence indication evident (i.e., caps and full stops). Some other basic punctuation used correctly (e.g. caps for proper nouns, commas in lists). 	<ul style="list-style-type: none"> Simple sentence indication evident (caps, full stops, question marks). Correct use of apostrophe for contraction Dialogue, where used, distinguished from prose by punctuation Errors do not interfere with comprehension. 	<ul style="list-style-type: none"> Consistent correct simple sentence indication (i.e., period, capital letter, question mark, exclamation). Mostly accurate use of colons, hyphen, apostrophe of possession when used. Dialogue punctuated accurately, though not necessarily conventionally. Largely conventional capitalisation of titles and proper nouns 	<ul style="list-style-type: none"> Largely consistent sentence indication Accurate use of apostrophes, colons, hyphens. Some success with embedded, parenthetical, and conditional phrases or clauses using commas, semi-colons. 	<ul style="list-style-type: none"> Accurate sentence indication for all types Largely accurate use of conventional punctuation within-complex sentences (comma, colon, semi-colon, hyphen, dash) Errors or omissions are few and do not impact on meaning
Spelling	<ul style="list-style-type: none"> Many HFW* (Lists 1–4) spelt correctly. Evidence of some knowledge of common spelling patterns and approximate phonetic spelling. 	<ul style="list-style-type: none"> Most HFW* (Lists 1–6) and common spelling patterns correct. Spelling shows some understanding of more complex spelling patterns. 	<ul style="list-style-type: none"> Few errors within HFW (Lists 1-7) and common spelling patterns evident Some success with multi-syllabic, irregular, or technical words 	<ul style="list-style-type: none"> Some success with complex multi-syllabic, irregular, or technical words. 	<ul style="list-style-type: none"> Mostly sustained control of complex multi-syllabic, irregular, or technical words. Errors or omissions are few and do not impact on meaning

Note: ignore the spelling of complex words copied directly from stimulus material. Evaluate spelling based on student’s own words.

Appendix 2. Training Materials for the Scoring of Writing Levels 2—6



asTtle:
*Assessment Tools for
Teaching & Learning*



●
● Markers Briefing
●

School of Education
The University of Auckland



The five KEY words

- ★ **Professional judgment**
- ★ **Consistency**
- ★ **Accuracy**
- ★ **Confidentiality**



Professional judgment

requires you to use your knowledge of students and the curriculum to check that...

- ★ **the examinee's writing is at the level described by the scoring rubric**
- ★ **the evidence is sufficient for a reliable judgment**
- ★ **other markers would make a comparable decision.**



Consistency



- ★ Principle of marking “through one set of eyes” – an extension of me!!
- ★ “*Can have your say, but might not have your way.*”
- ★ Stick to the rubric!



Overview



1. Work in teams of seven.
2. Mark guinea pig script(s) at start of each day. Confer and check. Hand in this script.
3. Mark first batch of 5 scripts noting any aspects of scripts that may require clarification.
4. Discuss scripts with team members - seek assistance if necessary.
5. Mark scripts according to the rubric.
6. At specified intervals, team leader will check mark a randomly selected script.
7. Repeat steps 5 and 6 until all scripts for that purpose are marked.



First task – be a student



- ★ **Read the task, and think about what students are asked to do.**
- ★ **Think like a student!**
- ★ **Check the rubric for the purpose, and note the key features of each level for each variable.**
- ★ **Discuss and clarify the task and rubric with your team.**





Scoring process



- ★ Read whole script. Consider options for invented spellings - focus on context and meaning. Know what the child is saying before you begin to score.
- ★ Think **FUNCTION** (What is the purpose here?).
- ★ Think **TASK** (What has the writer been asked to do?).
- ★ Read **CHARACTERISTICS** carefully in the light of the written text.



- ★ For each variable, select the best descriptor from LEVEL 2/3/4/5/6.
- ★ Check sample carefully for evidence to support your judgement.
- ★ Use the **Basic** and **Advanced** grades to augment your decision as to how well the **Proficient** descriptor fits the evidence in the writing.



Augmenting your decision



- ★ **Basic** (criteria at this level/category) Shows signs of these elements. Elements are evident in embryonic form.
- ★ **Proficient** (at this level/category) There is evidence that the student is controlling the criteria elements.
- ★ **Advanced** (at this level/category) Student is consistently meeting the criteria at this level. Little disconfirming evidence is found.



Being seduced?



- * **Think task:** Good (and lengthy) writing which does not address the question.
- * Lots of content but incidental rather than central to argument.
- * Interference from surface features (e.g. spelling, handwriting, punctuation).
- * Knowing or making inferences about the writer from information given.
- * Interesting language use that does not contribute to the task.



Marking the script



- * Use supplied **ultra violet** pen to mark.
- * Once you have determined level for variable, enter the numeral.
- * **Augment numeral with a sub-level:**
 - Advanced
 - Proficient
 - Basic.
- * Enter only **ONE** level/score for each variable. No answer to read – enter a carat ^ in each box.



- * **Diagonal line through unused lines on last page used.**
- * **Wrong level entered? Cross out error, write new level and **initial** the change.**
- * **Make score legible use plain numerals (2, 3, 4, 5, 6) and CAPITAL letters.**
(Note: use >6A for student who exceeds 6A, and <2B for student who does not meet 2B standard.)
- * **When script complete, enter your marker number at top right of cover.**



Check marking

- ★ **Quality assurance of process.**
- ★ **Done by the appointed check marker in each team.**
- ★ **Acknowledges accurate marking.**
- ★ **Picks up inconsistencies etc.**



Check marking - rationale

- ★ The challenge → accurately scoring student writing.
- ★ Three ways of checking this:
 - Consensus
 - Consistency
 - Measurement.



Consensus

- ★ Degree to which scorers reach same decision.
- ★ Measured two ways:
 - Percent exact agreement
 - Percent adjacent agreement (scorers are within plus or minus one score level).



Consistency



- ★ Degree to which pattern of high and low scores is similar.
- ★ Pattern may be the same, but have to allow for *harsh* and *lenient* markers.



Measurement



- ★ Degree to which scores are result of common scoring, with minimal error.
- ★ Takes into account the task, the scorer, error and the interaction of these.
- ★ Complex, but aims to establish degree to which assigned score represents a *true* score.



Check marking



- ★ When asked by check marker:
 - enter your marker code on top right of next script then score the script, BUT ...
 - enter script number & scores/levels on the Check Mark Record Sheet (∴ check marker is not influenced by your scoring)
 - pass this script to check marker with Record Sheet.



asTtle Writing Check Mark Record Sheet

Marker ID

When advised by the Check Marker:

1. write your Marker ID code on the front cover of the next script as usual
2. record your scores for that script on **this sheet only**, then hand the script to the check marker. Do not record the scores on the front cover.

Script Number	Autism	SB	SB	4P	4A	4A	4P	4P
195 354		SB	SB	4P	4A	4A	4P	4P



Check marking process



- * Use **red pen** (supplied).
- * **Read and score the script.**
- * **Enter levels in boxes on front cover.**
- * **Collect the marker's Check Mark Record Sheet.**
- * **"Flag" discrepancies – anything more than one sub-level.**
- * **Discuss variations, and try to resolve according to rubric.**
- * **Call supervisor if unresolved.**



Guinea pig scripts



- * Guinea pig (photocopied unmarked) script(s) at start of each day.
- * Common script for each table/purpose.
- * Score in usual way – scores in boxes, and marker number on front.
- * Discuss/justify your scores to other markers at table, but **do not change your assigned scores!**
- * Collected for statistical comparisons.