

**Assessment Tools for Teaching and Learning Technical Report # 49****Summary of Teacher Feedback from the Secondary School  
Calibration of asTTle Reading and Writing Assessments for  
Curriculum Levels 4 to 6**

Abstract: This report summarises the teacher feedback from the large-scale standardisation of over 10,000 reading and 7000 writing scripts completed by students in Years 8 to 12 in early 2004. This evaluative feedback is used to improve the quality of assessment materials in reading and writing before their publication in the forthcoming asTTle™ V4 software. Teachers identified as their most serious criticism of the asTTle™ test papers the mismatch in test paper difficulty with the ability of all students in their classes. The asTTle software allows teachers to customise test difficulty for the ability of their own students.



**Submitted by the Assessment Tools for Teaching and Learning team,**

**Auckland UniServices Ltd**

**University of Auckland**

**August 2004**

## **Summary of Teacher Feedback from the Secondary School Calibration of asTTle Reading and Writing Assessments for Curriculum Levels 4 to 6**

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 summarises the teacher feedback from the large-scale standardisation of over 10,000 reading and 7000 writing scripts completed by students in Years 8 to 12 in early 2004. This evaluative feedback is used to improve the quality of assessment materials in reading and writing before their publication in the forthcoming asTTle V4 software. Teachers identified as their most serious criticism of the asTTle test papers the mismatch in test paper difficulty with the ability of all students in their classes. The asTTle software allows teachers to customise test difficulty for the ability of their own students.

I would like to thank all the students and teachers across the country for completing the reading and writing tests.

A number of people were involved in developing this data. Dr Gavin Brown managed the process, and wrote this report. The data were transcribed and collated under the direction of Earl Irving and Kerry Sussex, with assistance from Sarah Cutforth.



John Hattie  
Project Director, asTTle™  
August, 2004

The bibliographic citation for this report is:

Brown, G.T.L., Irving, S.E., Hattie, J., Sussex, K., & Cutforth, S. (2004, August). *Summary of Teacher Feedback from the Secondary School Calibration of asTTle™ Reading and Writing Assessments for Curriculum Levels 4 to 6*. (asTTle Tech. Rep. 49). University of Auckland/Ministry of Education.

## Table of Contents

<b>Methodology</b> .....	4
<b>Results</b> .....	5
Content Appropriateness.....	5
Content Interest and Engagement .....	7
Teacher Instructions.....	8
Level of Difficulty .....	8
Time Allocation .....	9
General Comments.....	11
Concluding Comments.....	11
<b>References</b> .....	12

Formative evaluation is “designed, done, and intended to support the process of improvement” (Scriven, 1991, p. 20) and is normally carried out “by the staff of the originating institution” (Scriven, 1991, p. 22). Formative evaluation provides improvement-oriented feedback in response to three specific questions, which are (a) Where is the process or product now?, (b) What is the target for the process or product?, and (c) What steps are needed to reach the target? (Clarke, Timperley, & Hattie, 2003). It is critical in the development of a national teacher-controlled assessment tool that evaluative comment is obtained from the intended users. Research on teachers’ conceptions of assessment has shown that teachers are positive about obtaining improvement oriented evaluation of their teaching through student teaching but that they also consider externally mandated assessments as measuring only surface dimensions of learning (Brown, 2002). Thus, feedback from teachers who administer asTTle™ tasks is elicited in order to enhance teachers’ perception of the validity of the asTTle™ application. This report describes the formative evaluation conducted during the development of standardised assessment items in literacy for secondary schools and notes how the feedback is used to improve the quality of the assessment materials to ensure maximum impact on teachers’ use and acceptance of the new assessment tools.

The Assessment Tools for Teaching and Learning (asTTle™) project delivers a computer based set of tools for classroom, teacher-controlled assessment of student progress in literacy and numeracy at Levels 2 – 6 of the New Zealand curriculum in both English and Te Reo Maori. Specifically, this includes Reading, Writing, Mathematics, Panui, Tuhituhi, and Pangarau. Version

3 was released by the Ministry of Education to a sample set of secondary schools in early 2004.

asTTle™ Versions 2 and 3 are currently being used as a classroom assessment resource by teachers in both primary and secondary education.

### Methodology

New Zealand students in Years 8 to 12 completed tests of close reading and transactional or poetic writing targeted at curriculum levels 4—6 in March 2004. The reading test papers were of increasing difficulty as student year level increased; thus, Year 8 papers had largely Level 3 and 4 items, while Year 12 papers had mostly Level 6 items. The writing papers had tasks intended for use among secondary school students, with the majority of students being asked to complete ‘Analyse’ tasks in the senior secondary year groups. Note that the test forms as administered may be balanced in quite a different fashion to what an individual teacher would create using the asTTle™ application. asTTle™ permits teachers to customise the difficulty of the test to the perceived ability of the students being tested. For example, a teacher using asTTle™ can create a test with no or few hard items for a younger or less able group of students and vice versa; instead of having to use a standardised, centrally designed task as these trial papers were.

The trial papers were administered by teachers with their own class of students in a nationally representative sample of schools for the purpose of establishing New Zealand student performance norms. Table 1 shows that about 18,000 students completed the trial papers for reading and writing.

Table 1  
*Trial and Calibration Student Sample by Subject*

Subject	Sample Size	Number of Schools	Number of Teacher Responses
Reading	10,889	52	186
Writing	7,337	56	175
Total	18,226	108	361

Each teacher who administered the tests was asked to complete a form that structured feedback around key item development and usage characteristics. The key characteristics for which evaluative feedback was sought revolved around the quality of the items, specifically their

appropriateness in terms of interest or engagement, difficulty, length of time needed to complete, and around the quality of the instructions supplied with the test forms. In addition, teachers were asked for their own general comments, and were asked to summarise the nature of students' experience and evaluation of the materials. The teacher feedback was used to identify items and instructions that required adjustment in terms of language or length of time prior to publication in asTTle™ V4.

## Results

The total number of teachers possible to participate in the feedback was calculated as one teacher per 30 students. On that basis, approximately 610 responses could have been expected and 361 were received. This represents about 59% response rate by teachers. Nevertheless responses were received from at least one teacher at every school. Thus, readers can be confident that the feedback reported here represents a reasonably comprehensive body of opinion about the asTTle™ tests by currently practising secondary school English teachers.

Responses were in the nature of comments to prepared questions. The comments were generally coded using a “Yes”, “No”, “Both yes and no”, or “No answer”. The category “Yes” indicates a favourable or positive response to the question, a “No” an unfavourable or negative response to the question, and “Both yes and no” indicates a response that contains both positive and negative comments. “No answer” includes comments that were incapable of meaningful interpretation.

### *Content Appropriateness*

This section asked whether the content was appropriate for students' age and ability in the teacher's class (Table 2). Overall, teachers were positive about the content (89% of those answering the reading question and 66% of those answering the writing question). Positive comments included: *“this time of year, students should do this”*; *“worthwhile test”*; *“I thought the papers looked great, plenty of white space, well done”*; *“the students seem to have enjoyed the*

*content/material*"; "they liked the story best and the grandfather poem"; "Yes- it stretched some and others coped well"; and "yes- it required them to think and process information". In writing teachers' comments included: "this will provide us with useful feedback"; "challenged the students to think about what to write about"; "the idea of sustained writing is good to assess writing skills"; and "it was an excellent task".

Table 2  
*Appropriateness of content*

Subject	Number of Responses (Percent)				Total
	Yes	No	Both	No answer	
Reading	105 (56)	11 (6)	45 (24)	68 (37)	186 (100)
Writing	109 (62)	25 (14)	30 (17)	11 (6)	175 (100)

A number of themes arose from the various comments that teachers made about the content and have been used by the asTTle™ team to further refine the items. Many of the comments revealed an almost even split between those who thought the asTTle™ materials or items were appropriate and those who did not. These seemingly discrepant comments can be explained by the lack of teacher control over the materials in the tests they were sent; a feature that would be overcome in the asTTle™ application where they have more control over the content. For example, those who commented on the difficulty of the content were almost equally split over the reading materials and tasks being either too easy ( $n = 16$ ) or too hard ( $n = 13$ ). Likewise in writing, teachers were divided about the tasks being either too simple ( $n = 4$ ) or too hard ( $n = 8$ ). In a similar fashion, eight teachers commented that the content of three writing tasks was inappropriate requiring a level of prior knowledge that their classes did not have, while another five thought the tasks were ideal for their classes.

Some teachers commented that the language in the reading items (e.g., stereotypical, narrative, bias, assonance, allusion, cliché, colon, verse, fragment) was too difficult for the year level of their students ( $n = 7$ ). One teacher commented "the tests appeared to be written for pupils of a much higher ability and grasp of the English language.". It should be noted that other teachers

indicated that especially students from language other than English backgrounds would find the passages and items very difficult. Five teachers commented that the language of the writing prompts was too difficult. The feedback has been used to review items and prompts to enhance clarity of language. Nevertheless, the power of the asTTle™ application permits teachers to custom design a test for their own context; a significant advantage over other models of national assessment (Brown & Hattie, 2003).

*Content Interest and Engagement*

Teachers were generally positive about the interest and engagement of the content in the assessment items and tasks (Table 3), with five times as many saying yes than no for reading and nearly twice as many in writing. Positive comments about the reading tests included: *“the variety was appreciated by most students”*; *“information/context was varied - always good”*; *“content was interesting”*; *“some appreciated the variety offered. Some enjoyed the challenge of reading interesting passages”*; *“the layout was generally engaging and looked good”*; *“well set out - interesting and varied”*; *“students generally enjoyed the reading comprehension process”*; *“they were well received”*; and *“some commented they liked the fact that there was a variety, rather than just reading”*. In writing positive comments included: *“I was surprised, my whole low level class were engaged immediately and worked solidly for 30 minutes”*; *“students were very positive to whole process and are looking forward to getting their results”*; *“They enjoyed the topic”*; and *“yes- it was challenging for a number of my students”*.

Table 3  
*Content interesting and engaging*

Subject	Number of Responses (Percent)				Total
	Yes	No	Both	No answer	
Reading	87 (47)	16 (9)	59 (8)	24 (32)	186 (100)
Writing	77 (44)	42 (24)	43 (25)	13 (7)	175 (100)

Once again, a number of teachers indicated that the fit of the test to their own students was poor. The main symptoms included: students could not understand the reading passages ( $n = 6$ ), the

topics were not relevant or appropriate ( $n = 7$  reading;  $n = 14$  writing), and the topics were boring ( $n = 16$  reading,  $n = 15$  writing). Again this minority of teachers' comments can be understood by the lack of teacher control over the materials in the tests they were sent; a feature that would be overcome in the asTTle™ application.

### *Teacher Instructions*

Teachers were generally very positive about ease and usefulness of the teacher administration instructions for both reading and writing, with over three-quarters of teachers agreeing (Table 4). The negative comments revealed that there were too many instructions ( $n = 15$  reading,  $n = 9$  writing), the instructions were poorly formatted ( $n = 9$  reading), or they were difficult to understand ( $n = 6$  writing). It should be noted that, unlike many other standardised tests, this administration required teachers to break the reading tests into three 20 minute blocks and the writing tests required students to participate in a 10 minute brainstorming session. Further work will be done on improving the writing instructions before publication of asTTle™ V4. Note that the 20 minute block approach will not be a feature of asTTle™ V4.

Table 4  
*Teacher Instructions*

Subject	Number of Responses (Percent)				Total
	Yes	No	Both	No answer	
Reading	146 (78)	1 (1)	18 (10)	21 (11)	186 (100)
Writing	149 (85)	11 (6)	10 (6)	5 (3)	175 (100)

### *Level of Difficulty*

Teachers were asked whether the difficulty level of the items, passages, or tasks was appropriate for all students. Two thirds of the teachers believed the writing tests had appropriate difficulty, while a third of teachers were similarly positive about the reading instructions (Table 5).

Table 5  
*Appropriate Level of difficulty*

Subject	Number of Responses (Percent)				Total
	Yes	No	Both	No answer	
Reading	58 (31)	40 (22)	8 (4)	80 (43)	186 (100)
Writing	97 (55)	42 (24)	22 (13)	14 (8)	175 (100)

While the overall percentage of teachers giving positive responses was much lower for this particular question, it is important to note that this reflects that it is difficult to design a single test paper from a distance that is suitable for all students in any one teacher's class. The difficulty level of the test randomly assigned to a class, despite adjustment to fit the expected ability of each year level, was either too easy or too hard depending on the teacher's class.

The comments from those who provided them, revealed that the lack of fit was in both directions; that is, teachers were largely split over the reading tests being too hard ( $n = 33$ ) or too easy ( $n = 37$ ). Illuminating comments for the former position included: *"I have a low stream class (lots of ESOL) who didn't cope"* and *"for many it proved too difficult, but then I had to take into account the alternative class and level of ability."* Comments from teachers who believed the tests were too easy included: *"the students raced through each of the sections, so it may have been a bit easy"* and *"my mid-stream class seemed confident happy with the material"*. Again it is worth reiterating that in the released asTTle™ software, teachers will be able to customise assessments for the ability of not only whole classes but also for individual students.

#### *Time Allocation*

In the development and review processes around the Level 5 and 6 reading items it was noted by many participants that the reading passages were considerably longer than those used for Levels 2 to 4; especially for the connections content which required reading two or more texts. It was decided to allow two minutes per item to account for the large amount of reading. Further, the research literature indicated that low progress students manage time poorly in testing situations and as a result under-perform because they do not give themselves an opportunity to respond to items at

the end of tests. To redress this issue, the tests were broken into three 20 minute sections, each containing about 10 items, with instructions that students were not to go forward or return to any one section until or after the 20 minutes elapsed for that section. This will not be the requirement of the asTTle™ application, but did ensure that the trials provided valid information for all items.

Feedback from teachers about this novel timing method was elicited. Nearly a third of teachers ( $n = 58$ ) indicated, somewhere in the questionnaire, that the amount of time allowed to complete the reading items was too great; many teachers commented specifically that the 20 minute, 10 item sections were too long (i.e., “*students found it hard to refrain from moving onto the next section*”, “*can be done in 45 minutes*”). However, it was also commented that the “*a few ESOL students took the full time*”. Informal feedback from teachers in AimHi schools suggested that the 20 minute sectioning was seen as a positive feature. Thus, it would appear that the majority of students require less than 120 seconds per item, but some schools may prefer to allow that much time and make use of the sectioning technique.

In writing, teachers were instructed to allow students a five to ten minute brainstorming session before beginning writing, 30 minutes for individual writing, and then prompt a five minute individual review time before handing in the extended writing response to the task. A small number of teachers felt that the 30 minutes permitted for individual writing after the 10 minute brainstorming was insufficient. This concern was also two-edged; some felt that less able or motivated students needed more time to finish their writing, while others indicated that the highly motivated students wanted more time to extend their own work (e.g., “*students said they really enjoyed this task. They all wanted more than 30 minutes to write as many said they wanted to develop their story more and finish it*”). The use of the brainstorming was seen as a positive, as witnessed by these comments: “*planning part of the test was a good idea, helped students focus, and helped them discuss and make sense of it*”; “*mind map idea great*”; and “*many of them found it interesting once we had discussed it as a class*”.

These comments reinforce the need for standardised testing, as argued in Brown and Hattie, 2003, to be at the control of the teacher who can make, as was suggested by a couple of respondents, judicious and appropriate adjustments in procedures to fit their own environment (e.g., have the brainstorming on the day before the administration of individual writing exercise, or allow students to choose the task for the purpose that the student would complete).

### *General Comments*

When asked for general comments, a small number of teachers commented that it was difficult to get some students, and especially those in Year 12, motivated to take the asTTle™ tests seriously. This was so because the asTTle™ tests did not relate to their current unit of work or were not worth credit towards qualifications. Both of these concerns will, naturally, be best addressed by classroom teachers making voluntary, integrated use of asTTle™ tests within their teaching, as was intended by the design of the asTTle™ Project (Brown & Hattie, 2003). Some comments were made about students not having completed literature studies as part of their current year's work when the tests were administered in March. Teachers were concerned that student performance on the 'analyse' writing tasks would be reduced because of having to rely on previous year's work. Again, the publication of asTTle™ V4 will give teachers control over content, timing, and difficulty, mitigating these issues.

### *Concluding Comments*

The asTTle™ test items and materials in reading and writing were generally well received by teachers in terms of content, interest, and engagement, and for the instructions supplied to teachers. The feedback about the difficulty and time allowed for the test forms repeats the findings from the 2003 trial of mathematics items (Irving & Higginson, 2003) about teachers' concerns over the fit of the test forms to each teacher's class. However, with the asTTle™ software, teachers will be able to design and administer custom designed tests for the ability of classes and students and thus get around the fault of a centrally designed and administered assessment not fitting all environments

equally. The evaluative feedback will be used by the asTTle™ development team to adjust the length of time allotted to completing each item, to adjust teacher instructions, and most importantly in the design of the asTTle™ test creation process. Furthermore, the feedback indicates clearly that users can have confidence in asTTle™'s assessment material to engage and motivate students to show their true performance, provided teachers have administered tests of an appropriate difficulty level, given appropriate time to complete, and administered them at an appropriate point in the teaching programme.

### References

- Brown, G. T. L. (2002). *Teachers' Conceptions of Assessment*. Unpublished doctoral dissertation, University of Auckland, Auckland, NZ.
- Brown, G.T.L., & Hattie, J.A.C. (2003, July). *A national teacher-managed, curriculum-based assessment system*. asTTle Tech. Rep., 41, University of Auckland/Ministry of Education.
- Clarke, S., Timperley, H., & Hattie, J. (2003). *Unlocking formative assessment*. Auckland, NZ: Hodder Moa Beckett.
- Irving, S E. & Higginson R.M. (2003). *Improving asTTle for secondary school use: Teacher and student feedback*. asTTle Tech. Rep., 42. University of Auckland/Ministry of Education.
- Scriven, M. (1991). Beyond formative and summative evaluation. In M. W. McLaughlin & D. C. Phillips (Eds.), *Evaluation & education: At quarter century* (Vol. Part II, pp. 19–64). Chicago: NSSE.