Technical Report # 42

Improving asTTle for secondary school use: Teacher and student feedback

Abstract: This report summarises teacher and student feedback from the calibration (trials) of the asTTle v3 mathematics items, along with focus groups conducted with teachers concerning potential enhancements to the reports contained within asTTle. Analysis of student and teacher feedback from the asTTle V3 trial papers showed that the items were perceived as difficult, that students appreciated layout and design of the test papers, and that doing tests early in the year reduced opportunity to learn. Teachers favoured a networked, digital asTTle where students sit asTTle tests on computers with computer controlled scoring, data entry, and updating of a school wide database of results. They also understood and liked the revised data display mechanisms being proposed for asTTle V3.

Submitted by the Assessment Tools for Teaching and Learning team,

Auckland UniServices Ltd

University of Auckland

June 2003

Improving asTTle for secondary school use:

Teacher and Student Feedback

asTTle is funded by the Ministry of Education to Auckland Uniservices Ltd. at the University of Auckland to research and develop an assessment application for Reading, Writing, Mathematics, Pānui, Pāngarau, and Tuhituhi for Years 5-7 (Levels 2-4) 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 teacher and student feedback from the calibration (trials) of the asTTle v3 mathematics items, along with focus groups conducted with teachers concerning potential enhancements to the reports contained within asTTle. Analysis of student and teacher feedback from the asTTle V3 trial papers showed that the items were perceived as difficult, that students appreciated layout and design of the test papers, and that doing tests early in the year reduced opportunity to learn. Teachers favoured a networked, digital asTTle where students sit asTTle tests on computers with computer controlled scoring, data entry, and updating of a school wide database of results. They also understood and liked the revised data display mechanisms being proposed for asTTle V3.

I would like to take this opportunity to thank all the teachers, students, and schools involved in the trialling of Level 5 and 6 Mathematics items in 2003. Their responses to the survey have proved invaluable to the continuing development of the asTTle tool. A special thanks to the teachers who gave up their valuable time in order to participate in the focus groups.

Earl Irving and Raewyn Higginson of the asTTle team conducted and analysed the focus group responses. They also collated and analysed the teacher feedback reports completed as part of the administration of the asTTle tests in June 2003. Earl analysed the student evaluation questionnaire responses. They have successfully collaborated in the writing of this report. Dr Gavin Brown assisted in the writing of this report with editing and quality assurance.

Professor John Hattie

Project Director, asTTle

June 2003

The bibliographic citation for this report is:

Irving, S E. & Higginson R.M. (2003). *Improving asTTle for secondary school use: Teacher and student feedback.* asTTle Technical Report 42. University of Auckland/Ministry of Education.

Table of Contents

The ASsessment Tools for Teaching and LEarning (asTTle) project is in its third year of development. Currently over 90% of primary schools and a large number of secondary schools throughout New Zealand have requested copies of this new, innovative tool that delivers a computer based, teacher-controlled assessment of student progress in literacy and numeracy. asTTle V2 includes Reading, Writing, Mathematics, Pānui, Tuhituhi, and Pāngarau for Levels 2 – 4 of the New Zealand English and Māori curricula. In early 2004, a new version of asTTle will incorporate Levels 5 and 6 Mathematics and Pāngarau, while Reading, Writing, Pänui, and Tuhituhi Level 5 and 6 will follow in V4, due for release at the beginning of 2005.

The asTTle tool allows teachers to design a test by selecting the curriculum areas and levels of difficulty that they wish to assess. These selections are maximised by the asTTle tool to create a 40-minute pencil and paper test consisting of a mixture of open- and closed- response items. Once student responses and scores are entered into the asTTle tool teachers may select a range of reports that allow them to interpret student performance by reference to nationally representative norms, curriculum levels, and curriculum achievement objectives. Specifically, asTTle answers questions related to (a) how well are students doing compared to similar students, (b) how well are students doing on important achievement objectives, (c) how well students are doing compared to curriculum achievement levels, and (d) teaching resources that teachers are able to access that would assist in improving students' performance.

Six primary reports are generated and include the:

- **Console report**, which provides comparative /normative information about a group of students relative to similar students in the nation;
- **Individual learning pathway (ILP)**, which depicts detailed information on the strengths and weaknesses of individual students in the class;
- **Group learning pathway (GLP**), which provides information about the strengths and weaknesses of the class/group;
- **Curriculum levels report**, where student performance is analysed relative to the appropriate curriculum levels;
- **Tabular report**, which provides a 'mark book' summary of the performance of the class/group;
- **What Next report**, which links to the World Wide Web to provide teachers with high-quality teaching resources for use in the classroom relative to the curriculum levels and content by which asTTle tests are designed.

Because asTTle aims to inform learning and teaching practices of classroom teachers, it is important that the reports communicate clearly so that teachers are more likely to accurately interpret each report.

Formative evaluation is "typically done during the development or improvement of the programme… by the in-house staff … with the intent to improve" (Scriven, 1991, p. 168-9). It is in the best interest of both the developers and users of any national teacher-controlled assessment tool that evaluative comment is obtained from the intended users, and that the information is used to improve the quality of the assessment materials to ensure maximum impact on teachers' use and acceptance of the new assessment tools.

A significant feature in the development of the asTTle tool has been the extent of teacher consultation and input into the design of the items, software user interface, and the various reports. This development is on-going and the asTTle team is continually looking for ways in which to make asTTle more teacher-friendly and pertinent, especially in the area of test construction and the interpretation of reports. To date, in developing the Level 2-4 tool, this consultation has involved primary and intermediate school teachers. With the continued development of asTTle through to Levels 5 and 6, consultation with secondary teachers was seen as timely, especially following recent changes in student assessment brought about by the introduction of the National Certificate of Educational Achievement (NCEA) in 2002.

The three sections in this report discuss the teacher and student feedback asTTle has obtained through three separate studies. The first study reports feedback obtained from the Level 5 and 6 mathematics calibrations (trials) through surveys of

teachers and students. Two focus group studies comprised the second and third studies. These studies sought information on (a) test selection and reporting features of asTTle and their applicability to secondary teaching, and (b) improving the accuracy in teacher interpretation of one of graphical displays used in the reports.

Study 1: Feedback on Trial Papers

In June 2003, 823 mathematics items were trialled through a set of 24 different papers each containing 35-36 items. These items consisted of 598 new items written specifically for Levels 5 and 6 of the New Zealand Mathematics Curriculum, 48 items adopted from the AimHi project, and 35 link items from the existing V2 asTTle bank of items for Levels 2-4. The mix of these items in a paper varied across the Year levels, and 168 of the items appeared in two or more papers at the same or different levels. A total of 8,978 students in Years 8 to 12 from a nationally representative sample of 48 schools participated in the trials. The test papers were administered during normal mathematics class time.

Each teacher who administered trial papers with their class was invited to complete a feedback form. The potential pool of teacher respondents, estimated at one teacher per 25 students, would be 360. A total of 108 questionnaires were returned, representing approximately 30% of all teachers potentially involved. The feedback form had previously been used in asTTle trials and consisted of comments and ratings to prepared questions that asked about the appropriateness of the test items to class level, student responses to the test, teacher instructions, and also gave teachers an opportunity to provide general feedback (see Lavery & Brown, 2002 for a summary of the teacher feedback from the asTTle V1 and V2 developments).

In addition, a random selection of 1,197 students in Years 8, 9, and 10 were asked a series of questions about the test paper they had just completed. These questions covered whether they enjoyed the test, the difficulty of the items in the paper, the layout of the paper, how well they thought they had done on the test, and also invited their comments about the test. These student responses add direct feedback from students, in addition to the teachers' summaries of student reaction to the asTTle tests.

Results

Results for this study are reported by teacher and then student.

Teacher questionnaire

Appropriate Content

The first question asked teachers whether the content was appropriate for the age level and ability of the students. Table 1 shows the aggregated teachers' responses to this question which are categorised by whether they were positive, negative, or mixed. The teachers were divided on this issue; with equal proportions agreeing and disagreeing.

Table 1.
Teacher evaluation of content appropriateness

Category	Number	Percent
Yes	33	35
Mixed	26	28
No	34	37
Total	93	

The following comment helps elucidate the thrust of the negative comments:

> *For mixed ability classes this was too tough – to test whole ability range, ensure that there are easy questions so they don't give up.*

Level of Difficulty

Table 2 shows the aggregated responses to the question concerning the appropriateness of the level of difficulty for the year level of the students. Only one teacher indicated the test was too easy for the students. Almost half of the teachers thought that the tests were of an appropriate level of difficulty and two-fifths though the papers were too hard for the students. One teacher commented that the tests were higher than Level 4. As the tests were designed to trial items at Levels 5 and 6 (with link items to Levels 2-4 from an earlier asTTle version), this was reassuring.

Table 2.
Teacher evaluation of the appropriateness of level of difficulty

Category	Number	Percent
Easy	1	1
Appropriate	45	47
Hard	40	42
Total	96	

However, the persistent theme to the comments of the teachers who found the tests too hard was that the students had not had adequate opportunity to learn all the material presented in the trial forms. It is apparent from these comments that students

had not been taught all of the content objectives covered in the test, and for this reason the tests were too hard. Thirty-one of these teachers cited this as the reason that they thought the tests were too hard. Other comments included:

Needs to be more appropriate to year level
Too advanced for Year 10 work not covered
Content new to Year 10
Assessment items too difficult
Not fair, not covered topics.

Student Reaction

Teachers were asked to assess and report the reaction of the students in their class to the test papers. The responses to this were varied. In some cases, teachers tried to accurately reflect the voices of all students who expressed an opinion, while in others they gauged the reaction of the students while the students were answering the papers. Forty seven of the responses were negative in tone (*"Boring, pointless"* and *"7XX hated it*"), while twenty-six were positive in tone (*"They found it interesting"* and *"They seemed to enjoy it"*). A further thirty teachers wrote comments that were ambivalent (*"To begin with not good! As the time progressed they were happy with the challenge"*).

Teacher Instructions

In the interests of ensuring all asTTle tests are given in a similar way, asTTle provides a set of written administration guidelines. Teachers were asked to comment on the adequacy and clarity of those instructions. Seventy five of the teachers felt that the instructions were clear and appropriate, with thirteen teachers assessing them as complex and wordy.

Faulty Items

This question asked teachers to indicate whether any of the actual items in the test paper were faulty in terms of the written or visual material. Note that all items had been reviewed by one of three panels of secondary school mathematics teachers before going to trial. The teachers commented on fifty-one items. In most cases, this was to point out that the item covered content that had not yet been taught to the students in their class, that the content would not be covered in the school's syllabus for the given year level, or that the students needed drawing equipment to complete the answer. Nevertheless, nine items were found to have minor faults and were

rewritten to ensure that multiple-choice items had only one correct answer or to remove possible ambiguities.

General Comments

Teachers were given the opportunity to write open comments about the asTTle assessments. In general, they reinforced the opinions already expressed in response to the earlier questions. In keeping with the comment made above, opportunity to learn was frequently raised, with the suggestion that trials at the end of the school year would be more appropriate.

Student questionnaire

The student questionnaire asked students to rate ten statements using a positively packed rating scale (1=Strongly disagree to 6=Strongly agree, with two points expressing disagreement and four points for agreement). Table 3 shows the descriptive statistics for the student ratings. While the first seven statements were slightly favourably, the final three statements received very favourable ratings with each receiving average rating greater than 4.

Table 3.
Descriptive Statistics

Statement	Mean	S D	N	Missing

				N
I enjoyed doing this test.	2.98	1.40	1066	131
Doing these questions was enjoyable.	2.76	1.32	1061	136
The questions were harder than anything I have done at schools so far.				
Understanding these questions was easy.	2.93	1.35	1048	149
I am sure I did well on these questions.	2.70	1.34	1052	145
The questions were similar to the mathematics I have been doing in my class.	2.98	1.29	1045	152
I learned what I really knew in mathematics by doing these questions.	2.90	1.37	1028	169
I had enough space to work out my answers.	4.32	1.63	1039	158
I had all the equipment I needed to answer the questions.	4.32	1.75	1034	163
I liked all the white space around the questions.	4.30	1.62	1030	167

The ten items were subjected to factor analysis, using maximum likelihood extraction with direct oblimin rotation. The third statement (*The questions were harder than anything I have done at schools so far*) loaded weakly on all of the factors and was excluded from the analysis. The factor analysis revealed a three-factor structure to do with student enjoyment, layout of the test, and student confidence to do the questions. The pattern matrix is shown in Table 4 (loadings less than 0.30 have been omitted).

Table 4.
Student evaluation statements pattern matrix

	Factor

Statements	1	2	3
Doing these questions was enjoyable.	.941		
I enjoyed doing this test.	.773		
I had enough space to work out my answers.		.784	
I had all the equipment I needed to answer the questions.		.451	
I liked all the white space around the questions.		.422	
The questions were similar to the mathematics I have been doing in my class.			.671
Understanding these questions was easy.			.594
I learned what I really knew in mathematics by doing these questions.			.558
I am sure I did well on these questions.			.456
Mean Scale Score (*SD*)	2.87 (1.30)	4.30 (1.24)	2.90 (1.01)

The first factor describes student enjoyment in doing the questions and test, the second factor brings together statements referring to the layout and physical aspects of the test, while the third factor describes the ease with which students could do the items because of their similarity to that is done in classroom. Students moderately agree with the layout of the paper and the use of white space ($M = 4.30$), supporting asTTle's current layout design of pages. The other two factors have ratings not quite reaching slightly agree and would be best understood as students indicating that the tests were relatively difficult and not reflecting their class programs sufficiently.

Included in the questionnaire was an opportunity for students to write comments about the test. A total of 559 of the 1,197 students wrote comments, of which 129 simply said "No" they did not want to comment. The remaining 430 comments were aggregated into eight categories: - the test was easy; the test was hard; the test was both easy and hard (e.g. "some of it was easy but other parts were hard"); positive general comments; negative general comments; comments about themselves and their efforts on the test; suggestions about the test papers; and whether the tests covered material they had/had not learned (i.e., opportunity to learn). Table 5 shows the number of comments by year level by category.

Table 5.

Student comments

Category	Year 8	Year 9	Year 10	Total	Percent
Positive general	52	36	18	106	18.0
Negative general	14	19	31	64	10.9
Test was easy	8	1	0	9	1.5
Test was hard	58	96	42	196	33.3
Test was both easy/hard	12	9	6	27	4.6
Self comments	21	34	26	81	13.8
Opportunity to learn	9	15	5	29	4.9
Suggestions about test	24	23	29	76	12.9

A third of the students (33.3%) wrote that they found the tests hard. This was in keeping with the responses of approximately half of their teachers who thought that the tests were also too difficult for the students. Interestingly almost a half of those comments came from Year 9 students. Typical comments were:

> *These questions were really difficult*.
>
> *The way the questions were written were quite hard.*
>
> *I had trouble figuring out what they meant.*

Although very few students found the tests easy, nearly one in five comments were positive:

> *… it was an enjoyable test to do – sometime I would like to do it in the near future.*
>
> *The test was fun and educating for me, I had fun.*
>
> *Thank you very much for letting us do the test, it was great thank you.*

Year 8 students wrote the most positive comments with the emphasis on fun and enjoyment. Negative comments increased in frequency with year level. Approximately five percent of the students commented on the fact that questions in the tests were of material that they had not had the opportunity to learn, thus reinforcing their teachers' comments. The following were typical of the such comments:

> *… some of the questions are work that we have not covered.*
>
> *… some questions I didn't understand because we haven't done the subject yet*
>
> *… totally different than what we get taught at school*

The timing of the trials in June is a contributing factor. Trials at the end of the school year would overcome this, especially in mathematics where the opportunity to learn the content of the subject is critical to the students ability to be able to respond correctly and with the confidence that they have done well. However, this would mean that a new release of asTTle could not be ready before the middle of the following year.

It is interesting to note that almost one student in seven wrote a comment that reflected their own efforts on the test or their insecurity in mathematics classes/tests. Some of the comments indicated that the students saw a test like this as an opportunity to see how they measured up.

> *If I don't do good on this test I will know it's because I didn't understand the questions.*
>
> *It made me realise what I knew and what I didn't know.*
>
> *I am not good at maths so I did really bad I would say.*
>
> *I am really bad at maths and I don't know whether I will ever improve.*
>
> *I took a guess with the ones I didn't know.*

Research has shown that mathematics anxiety impacts on student performance (Hembree, 1990; Ma, 1999), and this anxiety is apparent from some of these comments.

The student comments and ratings support the teachers' assessment of the test papers – they were difficult for the students, and the lack of exposure to the subject content of items meant that performance would be under-reported as a result of these tests. However, because asTTle uses item response theory mathematics and standard setting judgement to calibrate the difficulty items and ability of students, this limitation does not impact on the validity of the asTTle score and level calculations. Nevertheless, trialling in Term 4 of any year is more likely to reduce anxiety; though this must be weighed against the difficulty of getting access to secondary students working at curriculum level 6 in their preparation for end-of-year examinations.

Study 2: Uses, Barriers, & Improvements in Secondary School

The second study consisted of three focus groups using a guided discussion approach to investigate how asTTle might be used in secondary school contexts. Three focus groups of secondary school literacy and mathematics teachers (Table 6) participated. Note that only two Maori-medium teachers participated (FG3) and so this report is does not purport to represent the views of Maori-medium secondary teachers. Furthermore, all the teachers involved in these focus groups had had only a limited exposure to asTTle.

Table 6.
Focus groups participants and purpose

Focus group (FG)	Participants	Number
FG 1	Secondary Mathematics teachers	7
FG 2	Secondary Mathematics teachers	6
FG 3	Secondary Literacy teachers	5

With the imminent extension of asTTle into curriculum levels 5—6 and high school years 9—12 in V3, it was deemed important to consult with secondary school teachers regarding their potential use of asTTle. For the purpose of these focus groups the objectives were (a) to gather teacher feedback on their current or future uses of asTTle, (b) to ascertain any teacher concerns/ barriers when using asTTle, and (c) to elicit suggestions that would make asTTle more 'user friendly' for teachers. The discussions were structured around the use of asTTle, the test construction process, and the interpretation of asTTle reports (see Table 7 for questions used).

Table 7.
Focus Group Guided Discussion Questions

Category	Questions
General Use	How do teachers expect asTTle to be of use to them?
	When would you use it?
	What would you use it for?
	What would you expect to find out from it?
Reports	Do the reports convey the sort of information you want/need?
	What changes could be made to improve the reports you receive?
Test Construction	In constructing the tests, what do you prefer?
	Items randomly mixed/easier-to-harder, or
	More choice over items in test
Barriers to use	What barriers do you think there are to using asTTle? (e.g., time, marking, expectations, data input, relationship to NCEA, PD …)
	How could these be overcome?
Assessment	What do you think assessment is?
	What is its purpose?
	How do you use assessments?
	When does assessment occur?

Results

This section describes the feedback presented under three sub-headings: current or potential uses, barriers or concerns, and suggestions.

Current or Potential Uses

Teachers generally indicated that asTTle would probably be used as a diagnostic tool (at the beginning of the year to sort students into streamed classes or to identify students who needed additional programmes/assistance to lift achievement); and/or as a summative tool (at the end of the year, or in comparing cohorts etc.). Only one teacher out of the twenty participants mentioned that she might use it to inform her own teaching practice. The focus groups implicitly indicated assessment focuses on the student's learning (to demonstrate change/show improvement/identify for individual assistance). Local area cluster use of asTTle is already being implemented in one area. The data are being used to evaluate incoming students for tracking or streaming into secondary schools and to establish cluster-wide patterns of learning and teaching priority.

The reports, in particular the Individual Learning Pathway (ILP), were favoured by teachers. They found the ILP to be an excellent tool that would be useful when reporting to parent and in giving feedback to students. In particular, they liked that the report mentioned strengths as well as weaknesses. This feature was an obvious draw card to using asTTle as a preferred assessment tool, as it was mentioned

that other tests did not make the students' strengths explicit. Another reporting feature all participants favoured was the curriculum levels. This page was deemed useful for sorting students at the beginning of the year and for identifying students who required additional assistance.

Barriers or Concerns

Administration of asTTle is the area where the biggest barriers or concerns exist. Three issues were central to all three focus groups: (a) photocopying costs; (b) item selection, and (c) the time it takes to enter student results into the computer. With budgets preset at the beginning of the year, teachers were reluctant to photocopy asTTle tests in the formatted version presented. Many resorted to cutting and pasting to reduce use of paper, with some teachers questioning different sections of the test itself (cover page and attitudinal questions). This barrier is a similar one reported by primary school teachers when questioned about their use of asTTle (Ward, Hattie, & Brown, 2003). Reduction of paper used in tests (e.g., no cover sheet, option to exclude attitudinal data) was suggested by some. Digital delivery and administration of the tests was suggested as a possible solution to this barrier.

Item selection for teachers was a concern. For mathematic teachers the ability to select individual items was mentioned as they may only teach a limited range of achievement objectives from one content area at a time. Whereas English teachers were concerned that the possible repetition of common reading passages in two different asTTle reading tests may falsely inflate performance due to previous exposure. They wanted the ability to exclude repeated passages for post-testing purposes, regardless of asTTle's ability to minimize previously used questions. A further suggestion was made to have a 'hard copy' reference set of all the individual test items.

Teachers were also concerned at the time it took to enter the results into the computer. This was augmented by the fact that teachers had to enter multiple classes (unlike primary school teachers who teach many subjects to one group of students). asTTle's use of both numeric and alpha keys added to data entry time and a suggestion was made to use the numeric keypad only. Although, this would reduce some time, the potential to confuse multiple choice answers and scored values increases significantly. Teachers also requested the ability to print out the student scoring grid so that results could be written in by hand and then passed onto someone

else for data entry. Note that digital delivery and administration of asTTle would no doubt provide automatic scoring of all selected-response items.

Some concerns were expressed about the potential publication of 'league table' comparisons of schools and the potential of asTTle being used to unfairly evaluate teacher quality. Both of these concerns are addressed in the design and intent of asTTle as explained in Technical Report 41 (Brown & Hattie, 2003).

Teachers' knowledge, use of, and access to computers is somewhat problematic. Although the MoE does provide laptops to schools for teachers, it appears that not all secondary schools have taken up this option. For some teachers, asTTle results did not easily compare to those of other assessments. For example, some English teachers indicated there was a mis-match between asTTle curriculum level reading scores and those derived from PATs and their own judgments. Further investigation and communication about this issue is required. One teacher questioned the apparent over-emphasis in mathematics on skills, rather than mathematical processes.

Suggestions

Teachers favoured the development of a networked, digital asTTle where classes of students could sit an asTTle test simultaneously on computers with computer controlled scoring, data entry, and updating of a school wide database of results. At least one group indicated a preference for school-controlled local area network version rather than a public-access internet version on the basis that students may practice asTTle tests at home removing teaching and learning information from teachers. Even if full digital delivery of asTTle is not feasible, teachers wanted a networked database with the ability to globally change student year levels at the end of each year.

The focus groups indicated that links into NCEA were necessary as preparation for qualification was a major goal in secondary school. One group indicated that NCEA achievement objectives were already being introduced into year 10 and the Unit standards were being used in Year 9. By providing clear links into NCEA in particular, the profile (and thereby use) of asTTle in the secondary school may be raised.

Study 3: Graphical displays

Graphical displays are a powerful means of communicating data to an intended audience, as a means of simplifying information rich data sets. The graphs should enable the audience to make comparisons, and draw accurate and useful conclusions from the data. Through convention, the most commonly used graphical displays are bar graphs, pie graphs, trend lines, and pictographs, and where the data is more complex, tables of data are usually presented. However, the use of tables has been shown to inhibit analysis, and reduce the capacity to interpret the data in a meaningful way (Henry & Ginn, 1990)

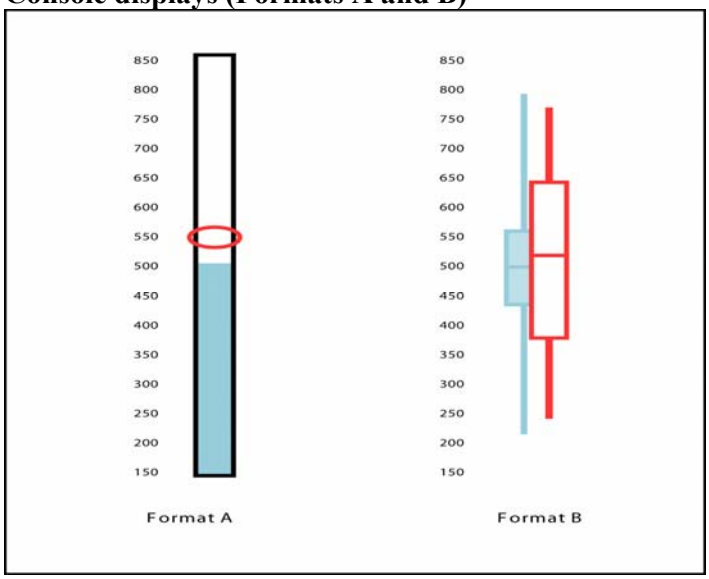
asTTle has adopted the use of graphical displays of data as a means of communicating assessment information to teachers (Brown, 2001; Meagher-Lundberg, 2001a, 2001b). These reports were developed in consultation with teachers to ensure that information was readily understood and interpretable. In a study conducted by Henry (1993), 277 teachers, principals, district superintendents, school board members, and print journalists evaluated two display formats, one of which was a modified box-and-whisker plot (Tukey, 1977). Henry found that the use of the box-and-whisker plot resulted in more accurate interpretations of the data than the presentation of the same data in tabular form.

The asTTle Console Report is designed to compare the tested students with the performance of similar students (e.g., year level, ethnicity, gender, and 'school like ours') derived from the asTTle national calibration. Initial feedback during asTTle V1 trial indicated that there was confusion over the meaning of the red ellipses on the Console Report. Consequently two focus groups of four primary and secondary teachers were invited to comment on the current Console Report, and a proposed box and whisker mechanism for reporting student performance. A semi-structured schedule of questions was used.

The two focus groups began by responding in writing to a set of read-aloud questions. The questions required interpretation of specific features of two different graphical formats – the existing bar graph and ellipse (Format A) and a proposed box-and-whisker plot (Format B) as shown in Figure 1. The current asTTle console report shows the average performance of the class/group using an ellipse, and the average norming group performance shown by shading in the bar. Of particular interest were the interpretations the teachers made relating to the ellipse in the barometer display,

which showed the mean score plus and minus one standard error of measurement for the group or class.

Figure 1.
Console displays (Formats A and B)



The box and whisker plot used in Format B was a variation on the standard five-point plot first described by Tukey (1977). It used the mean, plus/minus one standard deviation for the top and bottom of the box, and plus/minus two standard deviations for the ends of the whiskers. This means that the plot described the performance of approximately 95% of the group under consideration.

In the first focus group, participants were presented with a graph based on format A, asked comprehension questions, and given an opportunity to share their answers. Then, they were given a second example of Format A, asked questions, and shared their answers. Then, an example of Format B was presented, questions asked, and answers shared. Finally, a discussion comparing and contrasting the two report formats took place. The second focus group reversed the order of presentation of formats but followed the same sequence. The comprehension questions concentrated on interpretations that could be made about groups of average, below average, and above average students in the class, and how those interpretations might inform their teaching.

Results

The answer sheets were collected, and coded (Table 8) for an appropriate reading, recording, or interpretation of the format. As the questions were rephrased

following the first group, the results for the two groups are presented separately. A substantial majority of teachers were able to appropriately record the mean for the class and national norms using either format.

When asked to indicate the performance of the above average or below average students in the class using Format A, three of the four participants in the first group believed that they could inform their teaching of these groups from the information in the graph. Only one teacher indicated that it was not possible to tell from the information given; which is the correct interpretation. Participants in the second group were asked this question in a different way – firstly could they tell anything about the performance of the more/less able groups (Yes or No), and if so, what could they infer. This may have cued them to the correct response. As a consequence, all of them correctly responded that it was not possible to tell anything about the performance of either group within the class using Format A. It should be noted that the sharing of answers after the first example, seemed to have a marked effect on the accuracy of responses to the second example of the same graph.

Table 8.
Number of teachers answering correctly by focus group and format.

	FG1		FG2	
	Format A	Format B	Format A	Format B
Question	N	N	N	N
Appropriately records mean of class/group.	4	4	4	4
Appropriately records mean of norm group.	4	4	2	3
Appropriate interpretation of performance of class performance	3	4	3	4
Appropriately records measure of spread (standard deviation) of class/group.	NA	1	NA	0
Appropriate interpretation of performance of above average students in class.	1	4	4	4
Appropriate interpretation of performance of below average students in class.	1	4	4	4
Appropriate interpretation of second example.	3	4	3	3

Appropriate interpretation of the multiple year level	NA	3	NA	3

Following the presentation of the two examples for Format B, the teachers were shown a multiple year box and whisker and asked what they could infer, whether the students had made progress, and how they could tell. Three-quarters of the teachers were able to draw an appropriate conclusion concerning the different year levels, and whether students were making progress as they moved through the year levels.

One principal favoured Format A for reporting to the school board, as this provided a simple snapshot that could be readily understood and was quickly accessible, not requiring further explanation. All of the other participants (including another principal) favoured Format B, as they felt that there was more information in the plot that could help them in diagnosing, planning, and delivery of their teaching, as well as for strategic planning within the school. Teachers in the second group commented on their ability to group students using the additional information and teachers in the first group felt that the range of ability demonstrated in this way was more useful.

At the same time, participants commented that the box-and-whisker plot and the calibration scale were too far apart for an accurate reading to be easily obtained. They preferred for the calibrations scale to be closer to the plot. The participants in the second group disagreed that this distance distracted from the readability of the format. One participant liked the fact that both of the formats were not cluttered with gridlines.

Both formats had visual cues that were a source of confusion. The use of the ellipse on Format A was confusing, as the reader was unsure whether to read the top middle or bottom of the ellipse in order to obtain an accurate reading of the mean. The box and whisker plot in Format B was not constructed according to the convention that most teachers familiar with this type of plot might understand. The conventional wisdom uses five points – the top, bottom, median, and the two quartiles (or hinges) – but the use in this survey of the mean and standard deviations to construct the box, and the length of the whiskers would almost certainly lead to misinterpretations.

Both groups (including those teachers who have had prior exposure to asTTle) asked for an explanation of the scales shown. The length of the scale was commented on. The participants were unsure about where the scale started and where it finished, and this influenced how they interpreted the scores. In two cases, they expressed the scores as percentages in their written answers. They had assumed that the scale was calibrated from 0 to 1000, even though the scale displayed 100 at the bottom and 900 at the top. It was not generally understood that the scale was designed to record achievement on a scale calibrated such that the mean performance of Year 5—7 was set to 500 with a standard deviation of 100.

The issue of the use of colour was commented on, with a consensus that while the use of colour on the computer console was attractive, most teachers would use black and white when printing, and this was just as effective. However, if colour printing were to be used, then the brighter colour should be used for your own group, and not for the norm group.

When there are fewer than five students in a group for class, it is not possible to construct a box-and-whisker plot. The sheet showing multiple year levels had an example of this for Year 9. The data for the four students were shown as four numbers alongside the norm group box-and-whisker. Participants asked for more classes in Year 9 to be shown. However, it was not understood that this was a way of showing the performance of the four students in Year 9 alongside the norms for that level.

Discussion

There was one notable difference between the interpretations that the teachers made when using the two formats. In the first group, the teachers read too much into the Format A plot. The bar and ellipse only gave them information about the average performance of the class, yet they felt able to make inferences about how they would treat the more able and the less able students in the class. If teachers are using the information in this way, then we have reason to be concerned about the appropriate interpretation of this section of the Console Report. However, if they were to support this inference by reference to the one or more of other reports in asTTle (e.g., Individual Learning Pathways, Groups Learning Pathways, and Curriculum Levels), then there would be justification for such an interpretation.

The teachers were almost unanimous in their preference for Format B, as it contained more information on which to base sound decisions about the class and their teaching. As teachers they felt that this format was more useful. The sole exception to this consensus was a principal who intended to use the asTTle console report with the school board, and who felt that the present report was a simpler indicator for them to read and understand. However, teachers who were familiar with box-and-whisker plot expressed a preference for the conventional points in the data (top, bottom, median, and quartiles) to construct the plot used for the Console Report.

Conclusion

These three studies have given considerable feedback for asTTle development and implementation in secondary schools, some of which is being presently adopted or investigated. For example, the use of box-and-whisker reports has been implemented in the Console Report for asTTle V3. Specification of asTTle networking of data is being conducted, and investigations into relationship to NCEA and possible item specification mechanisms are being planned. These studies have also identified ongoing issues that would require significant changes to the current asTTle software; specifically the creation of a digitally delivered, fully-networked asTTle to resolve cost and time barriers. The issue of timing of trials remains open, and fuller communication with teachers about this issue is needed. Further evaluation with a functioning secondary school version of asTTle is being planned in conjunction with release of asTTle V3 to secondary schools in 2004.

References

Brown, G. T., & Hattie, J. A. (2003). *A National Teacher-Managed, Curriculum-Based Assessment System: Assessment Tools for Teaching & Learning (asTTle)* (Technical Report 41). Auckland, NZ: University of Auckland/Ministry of Education.

Brown, G. T. L. (2001). *Reporting assessment information to teachers: Report of Project asTTle outputs design.* (Technical Report 15). Auckland, NZ: University of Auckland, Project asTTle.

Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. *Journal for Research in Mathematics Education, 21*(1), 33-46.

Henry, G. T., & Ginn, V. (1990). *Improving evaluation practice through graphical analysis: An example using "wall chart" data.* Paper presented at the Annual Meeting of the American Evaluation Association, Washington.

Lavery, L., & Brown, G. T. L. (2002). *Overall Summary of Teacher Feedback from the Calibrations and Trials of the asTTle Reading, Writing, and Mathematics Assessments* (Technical Report 33). Auckland, NZ: Project asTTle, University of Auckland.

Ma, X. (1999). A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics. *Journal for Research in Mathematics Education, 30*(5), 520-540.

Meagher-Lundberg, P. (2001a). *Report on output reporting design: Focus group 2.* (Technical Report 10). Auckland, NZ: University of Auckland, Project asTTle.

Meagher-Lundberg, P. (2001b). *Report output reporting design: Focus group 2.* (Technical Report 10). Auckland, NZ: University of Auckland, Project asTTle.

Scriven, M. (1991). *Evaluation thesaurus*. Newbury Park: Sage Publications.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley Pub. Co.,.

Ward, L., Hattie, J. A., & Brown, G. T. L. (2003). *The evaluation of asTTle in schools. The power of professional development* (Technical Report 35). Auckland: University of Auckland.