The Evaluation of asTTle in Schools:

The Power of Professional Development

Submitted by the Assessment for Teaching and Learning team,

Auckland UniServices Ltd

University of Auckland

May 2003

The evaluation of asTTle in schools
The power of professional development

This report provides an evaluation of the implementation of asTTle (Reading and Writing) into 110 New Zealand schools from August 2002 – December 2002. 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.

Interviews and surveys were used to elicit teacher feedback on assessment, professional development, ICT, ability to accurately interpret asTTle reports, experience and ease at using asTTle software, and impact on teacher practice. A structural model of responses was able to clarify interpretation of feedback. The major messages were that professional development needs to be oriented most towards encouraging an attitude of the value of ICT in teaching and learning. By positively influencing the teachers' attitudes towards ICT there is a higher likelihood of correctly using asTTle as part of teaching and learning. It is noted, that many of these New Zealand teachers came to the asTTle application with already high levels of such attitudes to ICT as important in teaching and assessment. Further, the accuracy of reports is more a function of the level of impact, perceived quality of the reports, and belief that the results from asTTle will make a difference. This underlines the importance, as outlined in the Theory of Planned Behaviour, of perceived usefulness to subsequent use and correctness of use.

We especially acknowledge the leadership that Lorrae Ward has brought to this evaluation. Lorrae oversaw the development of the survey, completed all the telephone interviews, input and completed analyses of the qualitative data, and conducted the initial runs of all quantitative analysis. Her skills working with teachers and professional development personnel have been exemplary and considerably added to the quality of this report. She also completed the initial drafts of this report. John Hattie completed the final report, and refined the analyses that were needed. Gavin Brown oversaw the design of the asTTle test questionnaire items, and assisted with interpretations and comments on the report.

We also thank all the teachers and principals for readily providing the information, and we particularly welcomed the good news about the use and value of asTTle. The constructive comments about the weaknesses have been well received and used to improve the application. We hope the teachers and principals can see their recommendations for improvement in the future versions. We also thank the Assessment to Learn (AtoL) professional development teams for their excellence in teaching (as evidenced in this report), and their willingness to provide extensive feedback information.

John Hattie
Project Director, asTTle
June 2003

Table Of Contents

Introducing innovations into schools often follows a cycle of suspicion, eagerness, trialing, shelving, and ignoring. This is especially so when the innovation is a new voluntary assessment tool, and when this tool is markedly different from current assessment methods. Many researchers have documented the effects of such innovations as Gestetners, radios, television, computers, as well as policy innovations such as decentralization and merit pay for teachers. Most have been noted for their failure to be adopted and the reasons are aplenty. The current study evaluates the introduction of a new assessment tool and a different way of thinking about assessment into New Zealand primary schools. The model for the evaluation is built from the theory of reasoned action, and particularly emphasizes the importance of 'perceived usefulness' of the innovation as a major precursor to the use and adoption of the assessment tool.

The innovation is 'asTTle' (ASsessment Tools for Teaching and LEarning), which is a teacher controlled tool for creating tests in reading, writing, mathematics and the three Maori equivalents of pānui, tuhituhi and pāngarau, although only reading and writing were the foci of this evaluation. The application is installed onto any computer (home or school, PC or Mac), and the teacher can create tests, input and edit student records, and request scoring and reporting information. It includes national norms, comparisons to cohort groups, diagnostic information at the student and class level, information about levels of achievement (based on the New Zealand Curricula), and provides directions for future teaching based on the difficulty, strengths and/or deficiencies of the students. Unlike other available tests, it is tied directly to the New Zealand curricula, is norm and criterion referenced, can be used in a formative and summative manner, and emphasises interpretations relating to teaching and learning (more information at www.asttle.org.nz). The first version (V1) of the asTTle application (in Reading and Writing) was released to 110 schools in June 2002 and this paper outlines the evaluation of this trial.

The major claim underlying the evaluation is that the most powerful predictors of the use of innovations can be summarized by the Theory of Reasoned Action (Fishbein & Azjen, 1975). The most important aspects of this model are the notions of intent to use and adopt. This intent to use and adopt is a combination of attitudes toward performing the behaviour (including beliefs and evaluations of the expected outcomes) and subjective norms (other colleagues seeing value, approving of the innovation, and using asTTle). If a person *perceives that the use* of asTTle will be positive, then s/he will have a positive attitude toward performing that behaviour. More recently, Azjen and Madden (1986) introduced the Theory of Planned Behaviour (Figure 1), which added a third determinant of behavioural intention, perceived behavioural control (i.e. control beliefs and perceived power). Teachers' control beliefs are influenced by how difficult the barriers are perceived to be, as well as the perception of how successfully the individual can, or can not, perform the activity.

Figure 1.
Model of Reasoned and Planned Behaviour
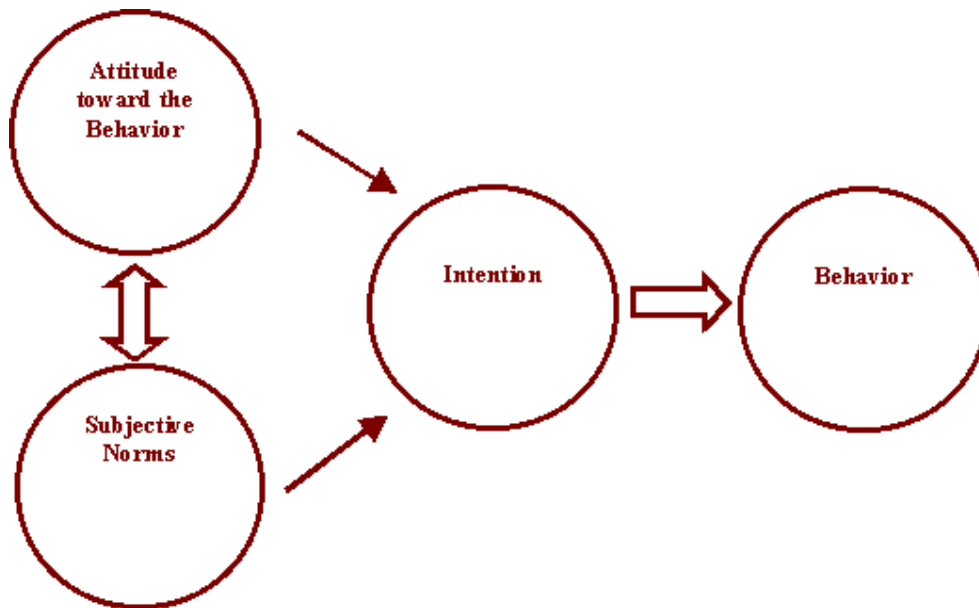
Among the more powerful subjective norms in schools are the beliefs that teachers hold about assessment, and the powerful defence of the status quo of teaching by most in school communities (principals, teachers and parents). As Tyack and Cuban (1995) have extensively documented, schools are "autonomous, buffered institutions. Educators have variously welcomed, improved, deflected, coopted, modified, and sabotaged outside efforts at reform" (p. 7), and they intensely like to protect what they consider is the "grammar of schooling". As teachers retain a fair degree of autonomy once the classroom door is closed,

> "They can, if they choose, comply only symbolically or fitfully
> or not at all with the mandates for change pressed on them by
> platoons of outside reformers. Or teachers could respond to reforms
> by hybridising them, blending the old and the new by selecting those
> parts that made their jobs more efficient or satisfying" (p. 9).

Tyack and Cuban (1995) identity two phases in education reform: (a) policy talk, and with respect to the current innovation this occurred during the late 1990's discussion on the Green Paper on Assessment and; (b) policy action through legislation and regulations and this occurred via the National Administration Guidelines which stipulate that schools must report to the Ministry on their success at enhancing student achievement, the commissioning of exemplars and the asTTle tools; and their implementation in schools, and which is the subject of this evaluation.

The Design of the Evaluation

Version 1 of asTTle (covering Reading and Writing) was introduced into 110 schools in August 2002. These schools responded to an advertisement in the Education Gazette. The schools were well spread across New Zealand, and represented all types of schools from small to large, rural to urban, low to high decile, and primary and

intermediate. Two-thirds of the schools received at least one day of professional development, conducted by the Assessment to Learn (AToL) facilitators, all of whom had been trained in the use of asTTle. Most participants at the professional development days provided responses about their experiences or were requested to email their reactions to the asTTle team. The AToL facilitators themselves provided further feedback. One applicant created a poem as part of the application:

> Yes! we want to participate
> In the asTTle pilot
> We will try our very best
> And give it all we got
>
> We fink a CD ROM
> Is a grand idea
> To help us wiv our learning
> And make literacy clear
>
> Our name is Maruia School
> Our address is a PO Box
> We've only got one senior teacher
> Who wears sandals, walk shorts 'n' sox
>
> His name is Peter Livingstone
> He teaches Yr 4 to 8
> I hope we can do this
> And are not tooo late!

With the reply

> We applaud your efforts to apply
> most others are dull, dreary & dry
> your choice of weeds is too good to hear
> but not yet hat, suit and tie?
>
> Keep up the wit, rhyme & verse
> for these words we indeed do thirst
> and your school's name may indeed
> get pushed right up to the first!

As part of the agreement to be involved, schools agreed to participate in at least one telephone interview (which was staggered over the 16 weeks of the trial with about seven conducted each week), and all users agreed to complete a survey. The survey consisted of three major parts. The first, common to all respondents, asked about precursors to using asTTle (e.g., adequacy of professional development, attitude to ICT), about the level and ease of use of asTTle, and about the value and worth of asTTle. The second consisted of a series of scales relating to attitudes to assessment, professional development, ICT, and asTTle. Third, every respondent was provided with at least 6 questions relating to the accurate interpretation of asTTle, and these were then converted to scores using a Rasch uni-dimensional IRT program that

equates for difficulty and places all items along a common scale (see Hambleton, Swaminathan & Rogers, 1992).

Three questionnaires were sent to each of the 110 schools and 176 were returned. As many of the schools were small (the average school size in NZ is only 150 students), the maximum number of possible respondents is somewhere between 250 and 330. Hence, the response rate was between 53% and 70%.

For each subset of items, the descriptive statistics are presented. A multivariate analysis of variance is used to ascertain whether there are differences in the means for principals and teachers, and for those who did and did not receive professional development. The sample size is not large enough to assess whether there are interactions between position and professional development.

Participant demographics

Teachers were the majority of respondents (46%), with 15% school principals, many of whom were also teachers). Deputy Principals (22%), and syndicate leaders (21%) made up the balance. Nearly three-quarters were in primary schools (74%) and one-quarter were in intermediate schools (23%). There was a spread of years teaching, although the majority had more than 11 years experience (Figure 2).

Figure 2. Number of participants for each period of time spent teaching
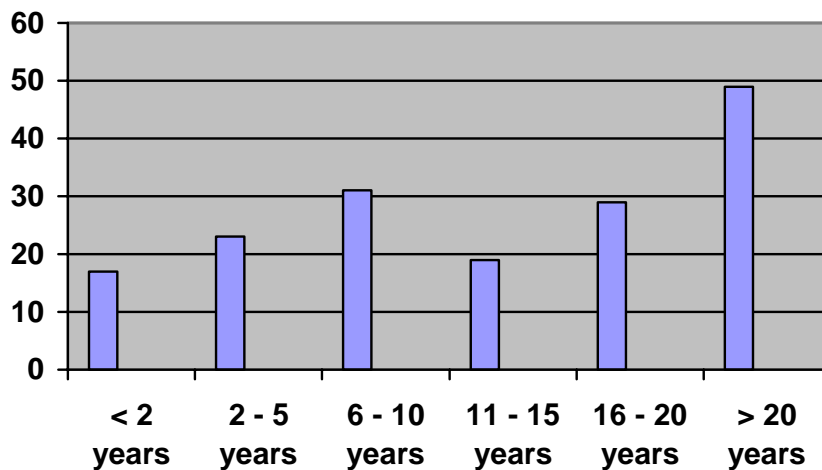
Installation and Use of asTTle

The first section of the questionnaire, administered to all participants ($n = 176$), included 17 questions relating to the use of asTTle by the participants.

Level of use of asTTle.
At the time of administration of the survey, asTTle had been in schools for about 10 weeks, and all participants had been requested to at least try a Reading or Writing test. Participants were firstly asked how often they had performed eight tasks using asTTle on a four-point scale (Table 1). All but five participants had used asTTle at least once

for one of the 8 tasks, 12% had used asTTle three or more times, and 80%+ had used asTTle for each task, with the noted exception that fewer had used the reports for future planning. Overall, this is a high usage given that only a few weeks were available.

Table 1.
Participant use of asTTle for performing key tasks.

Task	Not Yet	Once	2-3 times	> 3 times	% used
Entered student names and details for my class	15	56	22	7	85
Created a writing test	17	51	25	6	83
Administered a writing test to my class	21	65	13	2	79
Created a reading test	13	47	32	8	87
Administered a reading test to my class	16	71	11	2	84
Entered the scores for tests administered	15	50	28	7	85
Used the reports for future planning	38	33	23	5	62
Used the reports for reporting/discussion with others	21	40	27	12	79

Based on their total scores participants were placed into three different categories of use (8 – 15 = low, 16 – 23 = moderate and 24 – 32 = high with the splits based on 8 asTTle tasks times the mid points). The majority (52%,) were moderate users, 8% were high and 40% low users during this trial period.

There were no statistically significant differences over the eight level of use items between teachers and principals (Λ = .93, Mult. F = 1.42, df = 8.142, p =.196), but there were differences relating to professional development (Λ = .87, Mult. F = 2.59, df = 8.135, p =.012). For all items, the means were higher for those who received professional development (all effect-sizes > .30), indicating that the level of use of asTTle was greater for those who received some professional development.

Dissemination of information.
Participants were asked whether they had passed information from asTTle reports on to anyone else and if so whether this was done through discussion or in writing (Table 2). About half had used asTTle in discussions with students, two-thirds with syndicate or senior managers, and about a fifth with parents or with their school board. In the "other" category, 11 shared information with colleagues, one with the Educational Review Office, one with a principals' group and one with a curriculum review consultant.

Table 2.
Percentage of participants passing on information to identified recipients.

	Yes	Discussion	Writing	Both
Parents	23	17	6	1
Students	55	47	3	5
Syndicate	72	62	6	5
Senior management	78	63	9	6
School board	26	15	8	3
Other	11	8	2	0

There were statistically significant differences between principals and teachers over the five recipients of asTTle information ($\Lambda = .792$, Mult. $F = 7.30$, $df = 5.139$, $p = <.001$). Principals were more likely to present the information to parents, students, and school boards. There were also differences relating to professional development ($\Lambda = .891$, Mult. $F = 3.20$, $df = 5,131$, $p = .009$). Those who received professional development were more likely to present the information to senior management and the school board.

Ease of use of asTTle.
Participants were asked how easy they found five asTTle tasks using a six-point scale (1 = difficult, 2 = not easy, 3 = slightly easy, 4 = moderately easy, 5 = very easy, 6 = extremely easy - Table 3). The mean for all tasks lay between 4 (moderately easy) and 5 (very easy) implying that those participants found asTTle easy to use. The hardest task was entering student scores and the easiest was creating tests. The mean over all items was 4.32 ($sd = .86$), indicating that overall asTTle was "moderately to very easy" to use. Based on total scores, 52% found asTTle very easy to use, 44% found it slightly or moderately easy to use, and only 5% found asTTle hard to use.

Table 3.
Level of ease of performing asTTle tasks

Task	M	SD
Creating tests	4.94	0.91
Installing asTTle	4.29	1.41
Creating and interpreting reports	4.24	1.14
Entering student data	4.15	1.26
Entering student scores	4.00	1.34

There were no statistically significant differences over the ease of use of the five tasks between teachers and principals ($\Lambda = .920$, Mult. $F = 2.03$, $df = 5.117$, $p = < .080$), but there was between those who did or did not attend professional development ($\Lambda = .889$, Mult. $F = 2.82$, $df = 5,1113$, $p = .019$). Those who received professional development found it easier to install asTTle and enter student scores. There were no differences relating to creating tests, entering student data, or creating and interpreting reports.

Learning to use asTTle.
The trial was designed so that teachers in two-thirds of the schools received professional development and the others received none. In this sample, this proportion was mirrored: 28% indicated they had not, while 64% indicated they had received professional development (the remaining 8.5% did not answer the question). All participants, regardless of whether they had received professional development or not, were asked to state which of a number of given methods had helped them use asTTle. The responses were "did not use", "no" or "yes" (Table 4). The most frequently used type of assistance was being self-taught through trial and error. When this is added to the relatively high totals for use of the manual and the asTTle help function, then participants appear to have found working on asTTle themselves to be clearly the most beneficial method of learning how to use it. The least frequently used method was the ICT help desk, which was a free call to trained technicians/teachers (although it is noted that one in three teachers still used the Help Desk, with 42% of those using

it finding it helpful). A large percentage of participants (67%) received help informally from colleagues. The "other" category included help from a principals' cluster meeting, or from the technician in the school.

Table 4.
Percentage of participant responses that found each type of assistance useful

Type of assistance	% Yes
Self-taught through trial and error	76
Asked colleagues where necessary (informal)	64
External workshops/training sessions with an ATOL provider	63
Used the help function on asTTle	43
Read the manual	34
School workshops, training sessions with a colleague from school (formal)	34
Used the ICT help desk	13

**Adequacy of professional development.**
There were two sources of information relating to the adequacy of professional development: comments from those receiving the professional development, and from the survey, which includes those who did and did not receive the professional development.

Comments from professional development participants. There were nine teams offering professional development throughout NZ, to two-thirds of the trial schools. The Assessment to Learn (AToL) teams are Ministry of Education (MoE) funded and approved providers of teacher professional development with special responsibility not only for training on asTTle but also on other MoE assessment projects (e.g., ARB, Exemplars). The teams are based in the main population centres and serve the provincial areas surrounding the urban areas as well as the urban area itself. There are three in Auckland, one in Hamilton, one in Palmerston North, two in Wellington, one in Christchurch, and one in Dunedin. Evaluation comments were obtained from the participants through telephone surveys and were provided directly the AToL professional development providers. Overall, we received 832 comments from participants. Figure 3 presents the overall distribution of responses under the major categories, grouped by whether the comment was positive, neutral, negative or a suggestion for improvement.

All responses to "ease of use of asTTle" were positive, and included comments such as:
- Clearly allows them to see gaps in student learning and their attitudes to reading and writing - a plus;
- Has many excellent features and used effectively is going to assist our teachers in identifying student needs and planning focused programmes;
- See it as useful for sharing with parents provided parents understood it was only a snap shot using one tool.

Figure 3 .
Percentages of Feedback comments
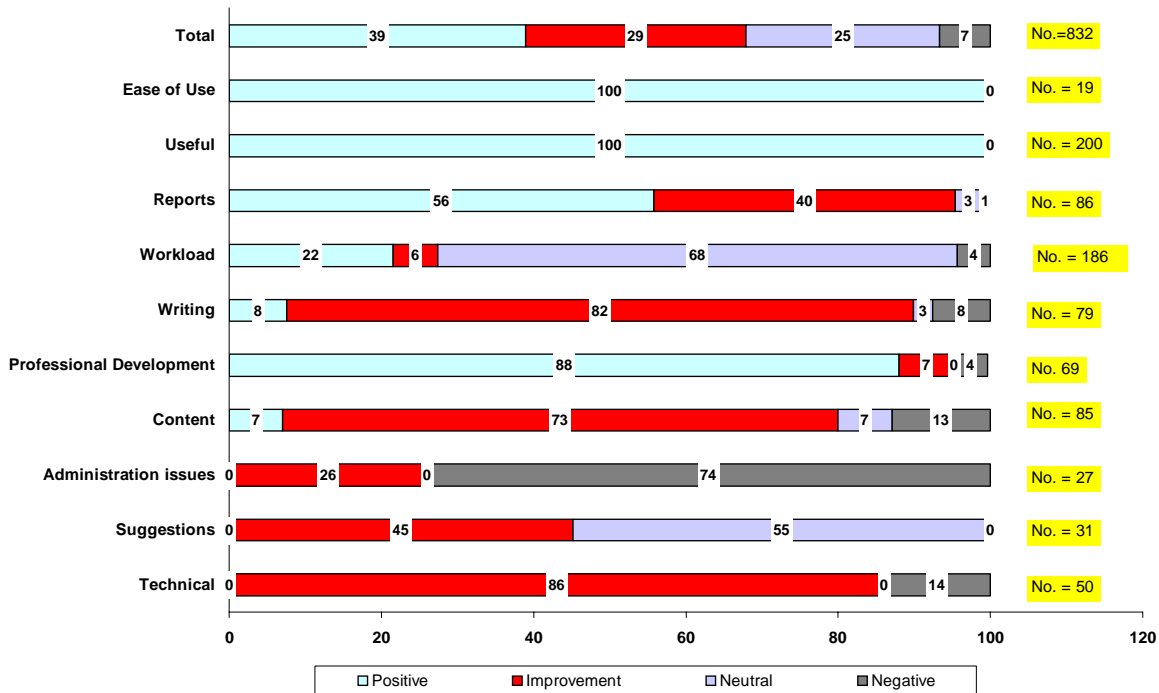
The majority of comments about the item contents reflected the importance of choosing items appropriate for the varying abilities of the students, for example: tests were horrific for lower achievers; would be good to have control over the content of surface features; several different genre tested in one test; how do we know what level each individual question (see also Lavery & Brown, 2002, which provides similar information from the evaluation of the trials of asTTle tests in the schools prior to release). Comments about Writing included: concerns regarding variation in marking between teachers (the manual outlines that teachers can attain inter-rater reliabilities of .8+ by following the scoring guide); grammar terminology needs more explanation; and the students liked writing tasks.

The majority of comments about Administration referred to the photocopying required. As the trials were in the later half of the school year, most of the teachers did not have sufficient resources for the photocopying or they saw the photocopying required as a wasting of resources. They commented: too many blank pages, and white space; cover wasteful each time; and will print double sided next time. There were some technical issues, such as: technical glitch when printing; cannot get Acrobat 5 on computer; need hour glass to show wait time; lack of computer memory; and thought had to use CD each time. Ten percent of the comments related to workload, such as: children did not finish writing in 40 minutes; long time to mark and enter scores; did not realise that you did not have to pre-score the closed items

(note: the contrasting response from the testing in schools documented in Lavery & Brown (2002), where the teachers and students welcomed the "white space" and the feel that the "tests" seemed more like classroom activities and not high stakes tests).

Just over half commented positively on the reports. Participants were: excited by amount of information in reports - ability to get quality comparative data - particularly useful in a rural school with limited numbers of students at year levels; teachers impressed with analysis of class performance and the what next connections, presented three tests to Board of Trustees who were very impressed. Most of the non-positive comments related to asking for more reports (e.g., there needs to be a simplified parent report; useful if it tracked individual asTTle results from year to year) or expressed some misunderstanding with the Group Pathways Report (which was redesigned in V2 on the basis of this feedback).

In the survey, those participants who received some form of professional development commented on the adequacy of the professional development they had received in the same areas (using a 6 point scale 1 = very inadequate, 2 = inadequate, 3 = slightly adequate, 4 = moderately adequate, 5 = mostly adequate and 6 = very adequate (Table 5). About 70% of participants who received any professional development considered it mostly to very adequate. The most adequate aspect of professional development was in the creation of the tests while the least in interpreting the test marks. Future professional development could add more attention to the marking of tests and interpreting the reports. These latter are the two areas that those who had no professional development also considered most important to be covered by a professional development course.

Table 5.
Participants' perception of the adequacy of professional development received

	M	SD
Creating tests	5.14	1.05
Interpreting reports	4.77	1.02
Entering school/class/student data	4.72	1.26
Entering student scores	4.60	1.30
Installing asTTle	4.59	1.45
Using the interpretations drawn from reports	4.45	1.13
Marking tests	4.40	1.20
Overall	4.67	

There were no statistically significant differences between principals and teachers, when asked about the adequacy of the professional development received ($\Lambda = .920$, Mult. $F = 1.19$, $df = 7,96$, $p = .315$).

Perceived benefits of Professional Development.
Those participants who had not received any professional development completed a question asking them how beneficial they believed it would have been to have had professional development in a number of areas. Table 6 summarises the means and standard deviations for this question. The scale used was 1 = detrimental, 2 = not beneficial, 3 = slightly beneficial, 4 = moderately beneficial, 5 = mostly beneficial, 6 = strongly beneficial.

Table 6.
Amount of perceived benefit if professional development had been provided

	M	SD
Using the interpretations drawn from reports	4.96	1.07
Interpreting reports	4.92	1.07
Marking tests	4.67	1.24
Entering student scores	4.46	1.37
Creating tests	4.42	1.37
Entering school/class/student data	4.41	1.43
Installing asTTle	4.20	1.65
Overall	4.58	

A total score was determined for the perceived level of benefit that would be gained through professional development by combining the scores for each participant. Three levels of benefit were determined using the score ranges of 7 – 18 (detrimental to not beneficial), 19 – 30 (slightly to moderately beneficial) and 31 – 42 (mostly to strongly beneficial). Across all items, these respondents claimed that it would have been moderately to mostly beneficial to have received some professional development. Over half of the participants who answered this question (55.3%) believed that professional development would have been mostly beneficial. The percentage of participants with scores in each range is shown in Figure 4.

Figure 4.
Percentage of participants in each level of perceived benefit of professional development
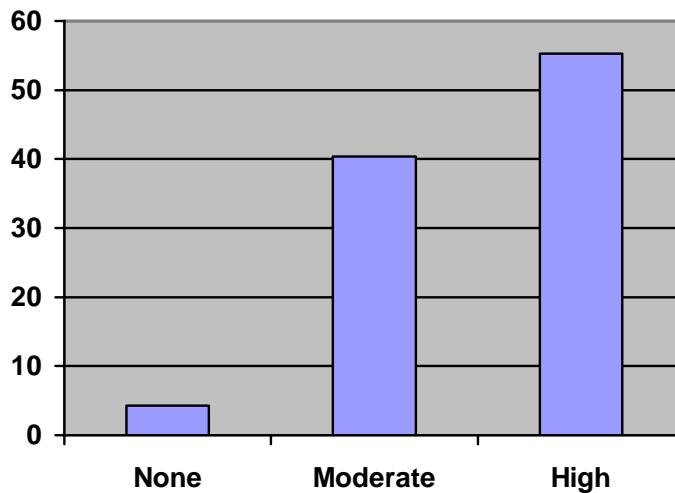
There were no statistically significant differences over the various tasks between teachers and principals ($\Lambda = .887$, Mult. $F = .71$, $df = 7.39$, $p = .662$).

Length of time spent on tasks.
On average participants spent between 6 and 30 minutes on individual asTTle tasks (Table 7). For a class of 30 the time taken to create a reading test, mark the scripts, enter the scores, create and interpret the reports would be five hours. The greatest time component relates to marking the scripts, which takes about 3.7 hours for reading and

5 hours for writing. (In Version 2 of asTTle, more assistance is provided for scoring, as many of the trial teachers did not realise that many of the items would be scored via the computer.)

Table 7.
Approximate time to fully administer a reading test to a class of 30.

Task	Range of time in minutes	Reading Average minutes	Writing Average minutes
Enter student details	16 – 60	38	38
Create reading test	6 - 30	18	7.5
Mark reading test	1 – 15 per student (x30)	225	540
Enter student scores	6 – 30 minutes	18	18
Create and interpret reports	6 – 30 minutes	18	18
TOTAL		**317**	**621.5**

There were no statistically significant differences in the time required for the tasks between teachers and principals ($\Lambda = .885$, Mult. $F = 1.44$, $df = 10,111$, $p = .171$), but there was between those who did or did not receive professional development ($\Lambda = .841$, Mult. $F = 1.95$, $df = 10.103$, $p = .047$). Those who received professional development took longer to familiarise themselves with asTTle, to enter student data for their class, and to mark one writing script.

Usefulness of asTTle

Three questions related to the usefulness and impact of asTTle. The first of these asked how much impact asTTle had had on a number of areas related to participants' professional lives. The second asked about the amount of new information gained through the use of asTTle while the third asked about the information gained from asTTle reports.

Level of impact.
Participants were asked how much impact asTTle had on a number of areas in their professional lives. The scale used was 1 = no impact, 2 = very little impact, 3 = some impact, 4 = moderate impact, 5 = a lot of impact, 6 = extensive impact. All of the means for this question were "moderate impact" (Table 8). There were some "other" impact areas identified by the teachers, such as reporting to the Ministry of Education, teacher associates, targeting group needs, and as a tool to confirm what was already known.

Table 8.
Level of impact asTTle has had on participants' professional lives

	M	SD
Teaching	4.14	1.14
The way you assess your students	4.09	1.13
Planning at a classroom level	4.08	1.20
Communicating with students	4.02	1.18
Communicating with senior management/school board	3.95	1.21
Communicating with syndicate	3.94	1.25
Planning at a syndicate level	3.71	1.35
Communicating with parents	3.84	1.27

There were statistically significant differences relating to the level of impact of asTTle between teachers and principals (Λ = .818, Mult. F = 4.00, df = 8, 144, p =<.001), but there were no differences between those who did or did not attend professional development (Λ = .958, Mult. F = .83, df = 8, 135, p =.582). Principals considered asTTle had more impact on the way they assessed students, communicating with their school board and with parents, and how teachers would teach.

Quantity of new information gained.
Participants were asked to consider how much more information asTTle had provided than they had had previously. A six point scale was used with 1 = none, 2 = very little, 3 = some, 4 = quite a lot, 5 = a lot and 6 = substantial (Table 9). asTTle provides "quite a lot" of new information about national comparisons, student strengths and weakness, and their learning needs.

Table 9.
Amount of new information participants have gained through the use of asTTle

	M	SD
Performance relative to others nationally	4.32	1.230
Student learning needs	4.06	1.17
Student weaknesses	3.93	1.16
Student strengths	3.87	1.16
Appropriate resources for students	3.40	1.36
Achievement objectives for students	3.39	1.23
Student curriculum levels	3.34	1.33

There were no statistically significant differences relating to the amount of new information from asTTle between teachers and principals (Λ = .933, Mult. F = 1.55, df = 7,150, p =.155), nor between those who did or did not attend professional development (Λ = .912, Mult. F = 2.00, df = 7,144, p =.060).

Information from reports.
asTTle provides users with six different kinds of report and participants were asked to comment on how much information they had gained from each on a six point scale (1 = none, 2 = very little, 3 = some, 4 = quite a lot, 5 = a lot, 6 = substantial). All reports provide 'quite a lot' of information, particularly from the Individual Learning

Pathways (Table 10). There is little difference between the perceived information provided by each of the reports, as the range for these questions is small (0.45). The top three reports provide information relating to the individual student, and collectively these provide more information than those more related to the class or groups of students.

Table 10.
Amount of information provided by asTTle reports.

Report	M	SD
Individual Learning Pathways	4.14	1.28
Curriculum levels	3.80	1.29
Tabular report	3.80	1.30
What next	3.79	1.46
Group learning pathways	3.78	1.25
Console	3.69	1.34

Combining participants' scores for this question gave an overall score for the amount of information provided by all asTTle reports. Four participants scored the minimum of 6 (no information gained from any reports) while 3 scored the maximum of 36 (substantial information from all reports). The mean overall score was 22.98 ($sd =$ 6.49) suggesting that participants on the whole found the reports provided "quite a lot" of information. There was a significant correlation between the amount of information gained from the reports and the overall level of use of asTTle (r = .263, p = .004). The greater the use, the greater the value of asTTle; and the greater the value of asTTle, the greater the use.

There were statistically significant differences relating to the information in the six reports between teachers and principals (Λ = .891, Mult. F = 2.47, df = 6,121, p =.027), as well as between those who did or did not attend professional development (Λ = .900, Mult. F = 2.19, df = 6,118, p =.049). Principals considered there was more information gained from the Individual Learning Pathways, the Console, and the Tabular Report, and there were no differences between teachers and principals on curriculum levels, what next, and group learning pathways. Those who attended professional development saw more value in the Individual Learning Pathways, curriculum levels, tabular report and group learning pathways. Attending professional development increased the perceived value of most of the reports.

Strengths and Weaknesses of asTTle

Participants were asked to list up to five strengths or advantages that asTTle offers as well five weaknesses, disadvantages, or suggested improvements. A total of 996 responses were given of which 545 were strengths (55%) and 451 weaknesses (45%). There was an average of 3 positive and 2 negative comments, and participants tended to offer both sorts of comments (r = .50). These responses were coded into 13 major categories, which were used for both weaknesses and strengths. Table 11 displays both the percentage of total responses for strengths within each category and for weaknesses. The range of responses for weaknesses was much more varied in content than for strengths because of the more specific and individual nature of many of the responses.

Table 11.
Percentage of responses for strengths and weaknesses within each category

Category	Strengths	Weaknesses
Quality & type of information provided	32.84	4.65
Ease of use	13.58	32.37
Comparative analysis possible	12.84	.67
Reporting & Planning tool	10.46	.67
Flexibility, adaptiveness to needs, tailoring tests	6.06	3.33
Type of assessment	5.87	.22
Features of asTTle	4.77	22.17
Content of tests	4.04	9.98
Other	3.85	12.86
Additional assessment tool	2.20	1.55
Consistency, reliability	1.83	3.77
Relevance to N.Z. curriculum	1.10	.22
Computer/technical issues	.00	4.88

Strengths.
The strengths primarily relate to the quality and type of information provided, the ease of use, that comparative analyses are possible, and its use as a reporting and planning tool. The most frequently occurring positive comments were the next steps report (which has also been the web page with the highest number of "hits", since asTTle was released to these schools), and the analysis of the strengths and weaknesses of students. Other themes included the ability to look at both individual and group results and information regarding levels.

The ability to make comparisons with other groups of students and schools was ranked as the second greatest strength of asTTle. Within this category the predominant type of comparison mentioned was being able to compare students nationally (61%), followed by comparison with other schools (17%).

Fourteen percent of the comments were related to ease of use. Within this category preparation of tests was seen as the easiest task while overall ease of use ranked second. The marking of tests was also seen as relatively easy, and other features included administering tests and reporting.

The flexibility of asTTle as a category includes all comments related to the ability to tailor tests to meet the needs of students and the amount of control teachers have over the tests they create. There were 6% of the positive comments regarding the flexibility of asTTle. The majority of comments in this category were related to the ability to design a test that met the needs of individual students, a class or a school. Also seen as a strength was the ability to assess specific objectives. Teachers liked being in control of the assessment.

asTTle was also perceived to be useful as a planning tool. The comments in this category were all based around the idea of using asTTle to inform teaching in some way whether through long term planning or more immediate planning as to the next step to take. asTTle is seen as providing a direction for teaching and a means of focussing teaching and learning.

The strength of asTTle for reporting included all comments that specified a particular type of report (the most noted were the "What Next" report followed by the "Learning Pathways"). Similarly, the extent to which asTTle was able to provide individual student data, the efficiency of use in terms of the speed with which things could be done, the level of focus possible, and the ability of asTTle to decrease a teachers' workload were also often noted.

Weaknesses.
The total number of negative comments (weaknesses or disadvantages) was less than that for strengths and advantages. The category in which asTTle was seen as having the most weaknesses is in the efficiency of use (Figure 5). All of the comments in this category were related to the time taken to perform tasks. Marking was seen as the most time consuming task ($n = 38$) with writing being mentioned more than Reading. Data entry was also a major concern with 35 comments related to the time taken to enter data. In V3 it is possible for various student management systems to provide a link directly into asTTle, which will help reduce this concern.

Figure 5.
Number of responses for weaknesses within each category
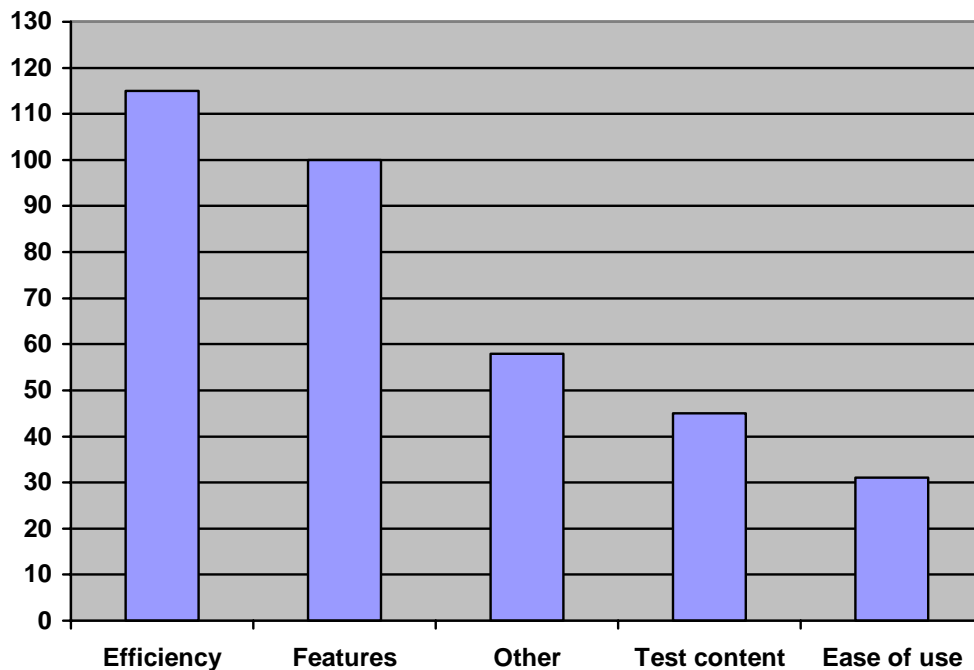
The second ranked category, 'Features of asTTle' is a very broad category encompassing a number of general comments related to the way in which asTTle has been developed and how it is used. The most frequently mentioned weaknesses concerned the amount of paper required to administer asTTle tests both from a cost perspective and also because there was perceived to be a lot of wastage ($n = 41$). The next most common group of comments were related to data transfer in some way ($n = 14$). For example, teachers wanted to be able to easily transfer data between computers either at school or between home and school (note: they can do this, but need to be more aware of this feature, and V2 has markedly improved this feature).

They also wanted asTTle to operate on a network (this was not part of the original brief, but may be added in future versions), and with the inputting of data (including the need to shift between a, b, c, and 1, 2, 3 operationalisation of the arrow keys in data entry, and labels on the multiple choice options).

Another composite category, "other", was ranked next ($n = 58$). This is because of the specific nature of many of the comments made in this area and the level to which they reflect personal biases or concerns. Comments related to lack of support (although AToL and a Helpdesk were available), request for reading ages, and more exemplars in the marking of writing.

Test content was ranked fourth ($n = 45$) with 21 of the comments relating to a lack of variety in the questions. Also within this group were a number of comments regarding the repetition of questions in different tests. V2 has corrected the impression that no item will be repeated. The application endeavours to create a *different test* (and there may be some repeat items), and the heuristic has been markedly improved for the choosing of items.

The only other category with more than 30 responses was "ease of use", including difficulties using a computer, the entering of data, the installation of asTTle, marking, and some of the language used in the asTTle writing progress indicators (e.g., tangential).

Other weaknesses related to technical concerns ($n = 22$), consistency ($n = 17$), and flexibility ($n = 15$). Within the technical category there were comments related to problems teachers had in operating asTTle on their computer such as the loss of data, problems with software, and the inability of school networked computers to run the program. The consistency category primarily included comments about the subjectivity of marking tests and interpreting results, while the concerns within the flexibility category were to do with the inability to specify precisely what teachers wanted in a test and the number of items in a test that not been specifically requested.

Overall comments
All participants were invited to make any further general comments. The majority were very positive, including:

- Principal and senior staff need to drive the implementation otherwise it becomes an 'add on' to already comprehensive/ad hoc assessment requirements;
- A comprehensive programme that provides a true analysis of individual children's strengths and needs. I really enjoyed interpreting the results and knowing exactly where student's strengths-weaknesses are. Fantastic to include the "where to next" help also;
- Awesome for future planning, teaching;
- asTTle has huge potential for planning and reporting by providing a benchmark comparison nationwide;
- Very affirming as it linked as far as levels go to other assessment tools; Confirmed suspicions & gut feelings about some children, reinforced findings of skills that are needed;
- Look forward to showing results with parents e.g. interviews, Very user friendly, Awesome links for 'What Next';

- Found training we received excellent, focused and very relevant;
- Would like to take the opportunity to say that I have really enjoyed being part of the trial;
- I have gained more individual information about my children in a short time than I would have over two terms, particularly in relation to their levels of thinking;
- I found that the full day professional development was of tremendous use and made it so easy to use within the school;
- The reports are clear and simple, and it is good to be able to compare the children's achievement with a variety of other schools etc.;
- More training on marking of writing, or at least far more exemplars;
- I especially like the idea of being able to compare my class's achievement with the norms for the whole country;
- The asTTle programme looks very exciting. I have now developed 2 reading tests and a writing test I have marked some of these It was very interesting marking the writing & assessing the curriculum levels achieved. I have used the tests for pre entry;
- Marking of writing scripts should become progressively easier as teachers build up expertise;
- This is an excellent resource - simplifies planning and reporting - will have an impact on measuring student achievement. Particularly like the fact it is not used to rank schools, etc.;
- We've been waiting for this for years. Well done!

Attitudes towards computers, asTTle, professional development, and assessment

Participants were asked about their attitudes towards computers, asTTle, professional development and assessment. It should be noted that not all participants answered all of the sections.

Computer confidence of participants.
Given that asTTle is a computer-based tool, one of the possible explanations for differential use of asTTle relates to the teachers' confidence in using a computer. Participants responded to an item about their confidence in using a computer on a six-point scale (1 = not at all confident, 2 = not very confident, 3 = slightly confident, 4 = moderately confident, 5 = very confident, 6 = extremely confident). The majority of participants (90%) felt at least moderately confident about their computer use while 28% were extremely confident (Figure 6).

Figure 6.
Number of participants at each level of confidence
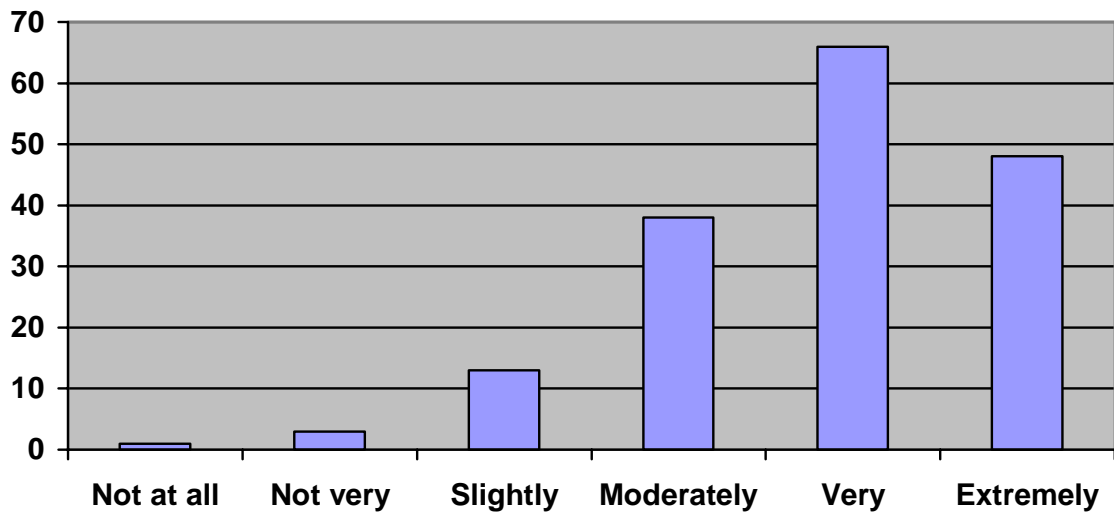
There were no statistically significant differences between principals and teachers relating to the confidence in using computers (Mult. $F = 2.47$, $df = 1,167$, $p = .118$), nor between those who did or did not attend professional development (Mult. $F = .50$, $df = 1,158$, $p = .479$).

Attitudes to ICT.
Participants were asked to indicate their level of agreement (strongly disagree to strongly agree) on 20 statements related to the use of computers ($n = 165$). There was a single factor underlying the relationships between these items, and the estimate of reliability was high (alpha = .96). Table 12 displays the means and standard deviations for each statement.

Table 12.
**Participants mean levels of agreement with ICT statements**

		M	*SD*
1.	A computer is a useful tool for my work	5.35	0.88
2.	To me the computer is a reliable learning tool	5.04	0.96
9.	I find it useful to have a computer handy when I am studying or working	5.02	1.26
15.	Nowadays I can hardly imagine working without a computer	4.96	1.21
5.	I welcome being able to use computers more in my work	4.88	1.10
20.	For a substantial part of my job, the computer is a useful tool	4.78	1.22
3.	When I work with a computer, I feel that I can get the computer to do what I want	4.75	1.17
6.	I can do tasks more smoothly with a computer than without one	4.70	1.24
14.	Many tasks can be done more easily and quickly using a computer	4.60	1.15
11.	Overall, I think the growth in use of ICT in our school has been worth it in terms of improved administrative efficiency	4.44	1.39
12.	Most of the technology I have seen could improve my students' learning	4.33	1.29
17.	Overall, I think the growth in use of ICT in our school has been worth the investment in terms of enhanced student learning	4.30	1.21
4.	These new technologies will substantially alter the way I teach	4.29	1.17
7.	Most of the technology I have seen could improve my ability to teach	4.11	1.25
10.	Working with the computer is never frustrating because I do understand the machine	3.67	1.72
18.	When I use the computer for work I am never concerned that it might break down	3.64	1.53
16.	I am able to understand computer problems that don't make sense to other people	3.63	1.68
8.	If I have computer problems while I am working I feel confident I can fix them	3.60	1.55
13.	I always understand what my computer is doing	3.51	1.51
19.	Computer technologies are easy to learn with relatively little support from outside	3.39	1.38

The overall attitude is clearly positive with mean of 4.35 (moderate agreement) and there is much support for the notion that computers are useful in teaching and learning. The items with lower means were more concerned with the ability to cope with the more technical aspects of computers.

A total score for attitude to ICT was calculated by combining individual scores for each statement. The higher the score the more positive the respondent's attitude to ICT as part of teaching and learning. The mean total score was 87.45 out of a potential maximum of 120 and a minimum of 20. Two participants scored the maximum of 120 while the minimum score for participants was 43 suggesting that none of the participants are overly negative about computers. These total scores were then grouped into three categories of low (20 - 53), moderate (54 - 86) and high (87 - 120). Figure 7 displays the percentage of respondents within each category and again reflects the very positive attitude these respondents have to ICT.

Figure 7.
Percentage of respondents at each level of positive attitude to ICT
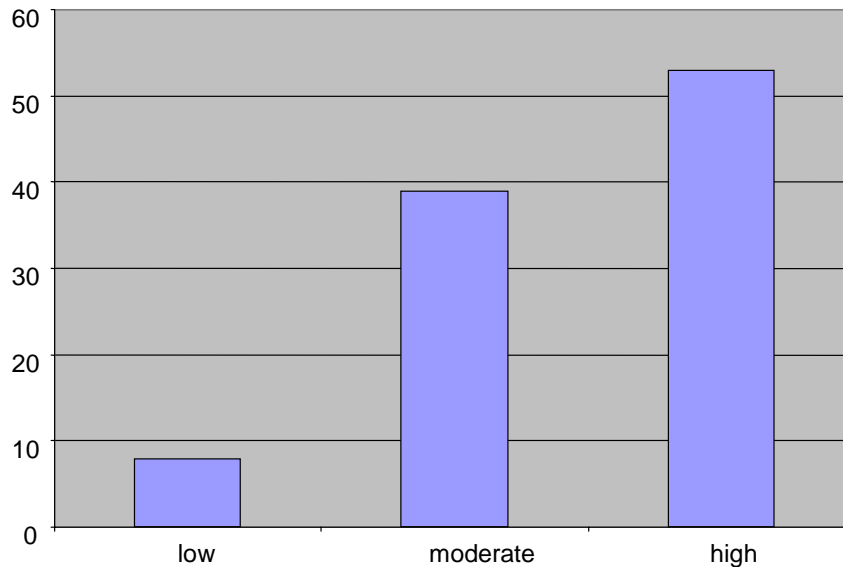
Principals were the most positive toward computers (*M*= 99), then deputy principals (*M* = 96), teachers (*M* = 85), and then syndicate leaders (*M* = 74). There were, however, no significant differences in attitude to ICT by years teaching, although the most positive attitudes were held by the most experienced teachers (> 20 years *M* = 90) and least experienced teachers (< 2 years, *M* = 99), with the most negative attitudes by the 11-15 year experienced teachers (*M* = 73). There were no statistically significant differences in the attitude to ICT between those who did or did not attend professional development (Λ = .586, Mult. *F* = .85, *df* = 20, 24, *p* =.643).

Attitudes to professional development /learning.
Participants were asked to show their level of agreement with 12 items related to learning how to use ICT applications such as asTTle (Table 13). A maximum likelihood factor analysis with oblimin rotation indicated that there were two factors (with three items loading on neither, relating to trying something new only when shown, need to be told via lecture first, and prefer someone outside the school to run training). The estimates of reliability were .60 for short practice-oriented training, and .79 for self-controlled, peer supported, informal learning.

Table 13.
Participants mean level of agreement with professional development statements

		Factor		
Practice Oriented Training	**M**	**SD**	**I**	**II**
2 It is best to have several short sessions of training with time in between to practice	4.54	1.28	1.01	-0.09
10 It is best to learn everything in several short sessions and not one long session	4.67	1.20	0.48	0.00
1 Professional development sessions are more beneficial if I have had time to try things out myself before hand	3.71	1.54	0.46	0.21
4 The best kinds of training/workshops are hands on involving real examples	5.46	0.84	0.32	0.14
Self-controlled, peer supported, informal learning	**M**	**SD**	**I**	**II**
7 I prefer to learn things informally from my colleagues rather than having formal workshops at a set time	2.78	1.03	-0.09	0.84
8 I prefer to learn new things at my own pace, when and where I want, using notes, instructions or manuals of some kind	2.88	1.37	0.10	0.82
3 I learn more if I teach myself rather than attending a workshop of some kind	2.91	1.26	0.16	0.69
9 It is better when a colleague from my school runs the professional development sessions than an external provider	2.91	1.11	0.07	0.62
11 I always read manuals or other instruction sheets before I start doing anything new	2.61	1.37	0.02	0.37

Correlation between the factors	I	1.00	
	II	.18	1.00

The most strongly agreed items related to a preference for practice oriented training sessions, while the methods with which participants somewhat disagreed related to preferences for lectures, or reading manuals. The participants also mostly agreed that several short sessions of training with time in between to practice was optimal, that professional development is more beneficial if there has been time to try things out before hand ($M = 3.71$), and it is better to have someone outside the school run training sessions ($M = 3.59$). Thus, the preference is for externally run, hands on, short sessions, with some pre-preparation exercise (Figure 8).

Figure 8.
Percentage of participants at each level of agreement for the idea that the best training is hands on involving real examples.
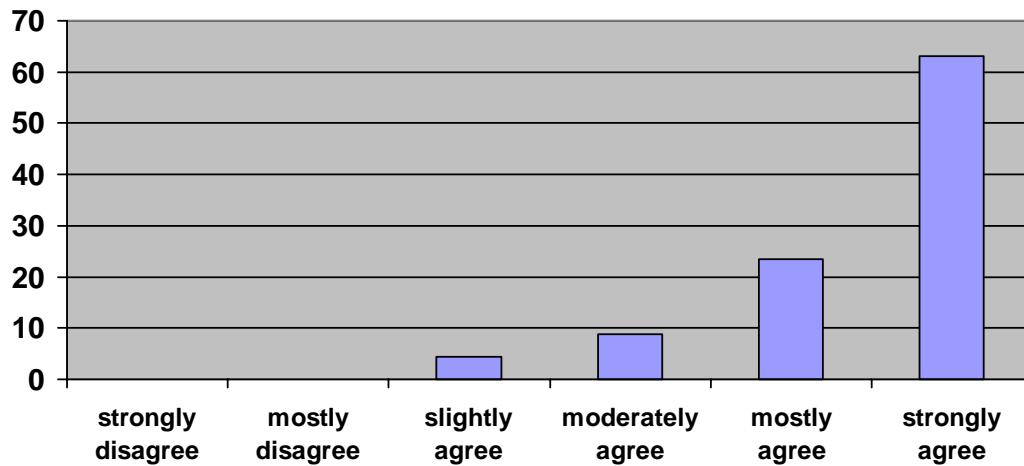
This can be related back to the earlier findings whereby the most frequently used method when learning how to use asTTle was self-teaching (82%), and 76% of these found this method the most helpful. Similarly 67% of participants stated that they had used informal help from colleagues with 64% finding this helpful.

There were no statistically significant differences relating to the information in the six reports between teachers and principals on both practice-oriented training (Λ = .957, Mult. F = .70, df = 4.62, p =.593) and on self-controlled learning (Λ = .915 Mult. F = 1.16, df = 10,124, p =.490). There were no differences between those who did or did not attend professional development for practice-oriented training (Λ = .947, Mult. F = .83, df = 4.59, p =.515) but there were differences on self-controlled learning (Λ = .761 Mult. F = 3.77, df = 5.60, p =.005). Those respondents who did not attend professional development favoured self-controlled learning (all effect-sizes > .50).

Attitudes to asTTle
All participants completed a series of 20 items concerning their attitudes to asTTle. Table 14 below shows the descriptive statistics for this section. A maximum likelihood factor analysis with oblimin rotation indicated that there were three factors (with one item, "It is possible to teach yourself how to use asTTle without any formal PD" loading on no factor). The estimates of reliability were high: asTTle is useful = .94, asTTle is easy to use = .79, and asTTle will make a difference = .90.

Table 14.
Participants mean level of agreement with statements

Item	Statements	*M*	SD	Factor Loadings		
				I	II	III
	asTTle is useful					
4	The tests created by asTTle are excellent assessment tools	4.61	1.09	.91	-.04	-.01
9	asTTle provides teachers with an accurate assessment of their class's achievement in literacy	4.64	1.11	.82	.03	.02
7	The content of the tests is excellent	4.42	1.12	.82	-.01	-.11
8	asTTle is a useful tool for classroom teachers wanting to plan ahead for their whole class	4.64	1.11	.70	.11	.09
5	asTTle is a useful tool at syndicate level for planning and discussion	4.53	1.17	.72	.10	.06
12	asTTle is a useful tool for planning to meet the needs of individual students	4.75	1.15	.66	.08	.18
10	asTTle is a useful tool for reporting to parents about their child's work	4.31	1.23	.54	.02	.35
3	asTTle is a useful tool for school management in overall school planning	4.58	1.14	.54	.12	.32
20	The levels of achievement within asTTle are easy to use and interpret	4.09	1.26	.49	-.03	.33
14	asTTle is a useful tool for talking to students about their own work	4.52	1.10	.40	.07	.40
	asTTle is easy to Use					
1	asTTle is easy to install	4.48	1.42	-.17	.94	.02
6	CD ROM is an excellent medium for distributing asTTle	5.00	1.21	.32	.57	-.15
2	asTTle is easy to use overall	4.34	1.22	.15	.48	.28
11	The Help function on asTTle is very useful	3.92	1.25	.22	.37	.22
	asTTle will make a difference					
16	asTTle will change the way we report to parents	3.64	1.19	-.07	.00	.92
15	asTTle will make a real difference to the way students are assessed in our school	4.09	1.16	.04	.08	.81
18	asTTle will make the assessment load of teachers a lot easier	3.44	1.48	.02	.02	.75
17	All teachers will find asTTle an invaluable tool	3.93	1.33	.38	-.09	.56
13	asTTle provides information that has not been readily available in the past	4.42	1.20	.24	.11	.48
	Correlation between the factors	I		1.00		
		II		.48	1.00	
		II		.67	.33	1.00

The items with the highest means included asTTle's usefulness for planning for individual students; the accuracy of assessment; a welcoming of the distribution via CD; planning for school management; and for talking to students about their own work. The items with the lowest means ("slightly agree") related to the Help function, reporting to parents, and reducing the assessment load.

There were no statistically significant differences relating to the attitudes to asTTle between teachers and principals on both Usefulness (Λ = .929, Mult. F = .95, df = 10,124, p =.490), Easy to Use (Λ = .964, Mult. F = .95, df = 4,103, p =.440), and Making a Difference (Λ = .962, Mult. F = 1.17, df = 5,146, p =.329). There were differences between those who did or did not attend professional development on Usefulness (Λ = .838, Mult. F = 2.39, df = 10,118, p =.017), but not on Easy to Use (Λ = .913, Mult. F = 2.36, df = 4,99, p =.058), nor on Making a Difference (Λ = .990, Mult. F = .37, df = 5,140, p =.927). Those who attended professional development considered asTTle had higher means on all items relating to Usefulness, particularly when using asTTle for talking to students about their own work, reporting to parents, and planning ahead for the whole class.

Attitude to assessment.
All participants also completed a series of 15 items concerning their attitudes to assessment (Table 15). A maximum likelihood factor analysis with oblimin rotation indicated that there were three factors (with two items, "Teachers conduct assessments but make little use of the results" and "Results should be treated cautiously because of measurement error" loading on neither factor). The estimates of reliability were high: Assessment for teaching = .77, School accountability = .80, and Student accountability = .72.

There were no statistically significant differences relating to the attitudes to assessment between teachers and principals on Assessment for Teaching (Λ = .919, Mult. F = .60, df = 5,34, p =.704), Accountability for Schools (Λ = .888, Mult. F = 1.47, df = 3,35, p =.240), and Accountability for Students (Λ = .902, Mult. F = .72, df = 5,33, p =.613). Similarly, there were no differences between those who did or did not attend professional development on Assessment for Teaching (Λ = .867, Mult. F = .98, df = 5,32, p =.443), Accountability for Schools (Λ = .954, Mult. F = .53, df = 3,33, p =.663), and Accountability for Students (Λ = .966, Mult. F = .22, df = 5,31, p =.952).

Table 15.
Participants mean level of agreement with statements

Item	Statements	M	SD	Factor Loadings I	II	III
	Assessment for Teaching					
5	Provides feedback to students about their performance	5.10	.83	.81	.16	-.05
10	Places student into categories	5.02	.85	.81	.00	.17
7	Is integrated with teaching practice	5.23	.8	.79	.25	-.15
12	Helps students improve their learning	4.80	.98	.52	-.04	.35
2	Forces teachers to teach in a way that is against their belief	4.93	.91	-.40	.39	.23
	School accountability					
1	Provides information on how well schools are doing	3.73	1.27	.02	.87	-.18
4	Is a good way to evaluate a school	3.00	1.15	.04	.81	.09
6	Shows the value schools add to student learning	3.93	1.1	.19	.62	.05
	Student Accountability					
11	Is checking off progress against achievement objectives	4.20	1.05	.21	.00	.76
9	Places students into categories	3.90	1.02	.22	-.05	.64
13	Is a way to determine how much students have learning from teaching	4.39	1.14	.35	.11	.54
8	Results are trustworthy	3.95	1.02	.28	-.01	.52
14	Is assigning a grade or level to student work	3.03	1.14	-.15	.24	.39
	Correlations between Factors			I	II	III
			I	1.00		
			II	-.05	1.00	
			III	.14	.23	1.00

Interpretation of the Reports

A series of three questionnaires were presented in a matrix sampling pattern (each teacher received two of the three questionnaires) that asked about the interpretation of the asTTle reports. For example, a console report (based on fictitious data) was presented and teachers asked "In which learning area did this class of students get the highest score?" or an Individual Learning Pathways Report was presented and teachers asked: What is the best way to understand the difference between Julie's ALS and the NZ reference group Year 5 mean?" Altogether there were 35 questions. The estimate of reliability of these items was .93, indicating that we can meaningful interpret these items and the total score.

The mean score was 20.84 or 60% (with standard deviation of 8.44). A stem-and-leaf diagram of these responses (Table 16) indicated that there is a slight negative skew (-

1.71) indicating that more teachers scored nearer the upper than lower end of this distribution

Table 16.
Stem and leaf of the Interpretation Correctness scores

Score	No. of teachers
0-5	15
6-10	7
11-15	12
16-20	24
21-25	76
26-30	24
31-35	18

 The means for each of the items is presented in the Table 17. The overall mean for the Console reports (.67) and What Next (.78) indicates a higher level of correct interpretations, but the means were lower for Individual Learning Pathways (ILP) (.51) and for Curriculum Levels (.57). Five of the items for Individual Learning Pathways were the lowest, particularly relating to Writing, and it is clear that more professional development for the interpretation of the Individual Learning Pathways for writing is needed. A major part of the misinterpretation was that too many teachers were interpreting the concepts and items in each of the cells (Gaps, Strengths, and Achieved) relating to the NZ or class norms and not relative to the student's average ability. The concept is "relative to this particular student's ability, here are the concepts and items that he/she should have got correct (i.e., items relatively higher than this student's average proficiency in writing or reading) but did not" (gaps), etc. Teachers, instead, were incorrectly interpreting the concepts as relative to the overall class or New Zealand norm group.

Table 17.
Means and standard deviations of the Interpretation Correctness Questions

No.	Report	Question	Mean	*SD*
1	Console	Ensuring identification of correct cohort group	.75	.37
2	Console	Interpreting relationships of scores to norm means	.60	.50
3	Console	Interpreting the relationship between scores and means	.71	.46
4	Console	Interpreting the relationship between scores and means	.62	.49
5	Console	Interpreting the relationship between scores and means	.73	.45
6	Console	Interpreting the relationship for classroom instruction	.24	.43
7	Console	Interpreting the relationship for classroom instruction	.82	.39
8	Console	Interpreting high and low console scores	.83	.38
9	Console	Interpreting high and low console scores	.55	.50
10	Console	Interpreting priorities for teaching	.87	.34
11	Console	Interpreting priorities for teaching	.70	.46
12	ILP	Interpreting the normative information	.71	.46
13	ILP	Interpreting the items relationships with objectives and quadrants	.69	.47
14	ILP	Interpreting the relationship between scores and means	.87	.34
15	ILP	Interpreting the relationship between scores and means	.71	.46
16	ILP	Interpreting the relationship for classroom instruction	.20	.40
17	ILP	Interpreting the relationship for classroom instruction	.22	.42
18	ILP	Interesting high and low ILP scores	.78	.42
19	ILP	Interpreting score-mean relationships	.85	.36
20	ILP	Interpreting meaning of quadrants	.86	.35
21	ILP	Interpreting relationships of scores to quadrants	.07	.26
22	ILP	Interpreting relationships of scores to quadrants	.07	.26
23	ILP	Understanding the quadrant and writing	.04	.20
24	Cur Levels	Interpreting least variation	.60	.49
25	Cur Levels	Interpreting central tendency	.81	.40
26	Cur Levels	Reasoning for interpreting central tendency	.67	.47
27	Cur Levels	Valid interpretation of scores to classroom instruction	.71	.46
28	Cur Levels	Valid interpretation of scores to classroom instruction	.40	.49
29	Cur Levels	Valid interpretation of scores to classroom instruction	.73	.45
30	Cur Levels	Valid interpretation of scores to classroom instruction	.56	.50
31	Cur Levels	Valid interpretation of scores to classroom instruction	.33	.47
32	Cur Levels	Valid interpretation of scores to classroom instruction	.31	.46
33	What Next	Interpreting score display	.83	.38
34	What Next	Interpreting score display	.75	.44
35	What Next	Interpreting for classroom instruction	.75	.44

There were no statistically significant differences in the interpretation score between teachers and principals (F=.02, df=1,172, p =.892), nor between those who did and did not attend professional development (F=2.41, df=1,163, p =.122).

The major correlates of interpreting the asTTle reports correctly were disseminating reports (r =.36), ease of use (.30), total use of asTTle (.22), rating of the six reports (.21), level of impact (.17), amount of new information (.17), usefulness (.17), and the reports will make a difference (.18). Those teachers with a conception of teaching relating to "assessment is powerful for teaching" had higher interpretation scores (r =.34), whereas those who had a conception of teaching as something relating to school accountability had the lowest interpretation scores (r =-.21). As argued by Brown (2002) professional development needs to attend to the conceptions of assessment held by teachers before introducing asTTle, because those who see

assessment as about school accountability rather than improvement of teaching are most likely to not accurately interpret, or attend to the information in the reports.

Those who most correctly interpreted asTTle made more positive (.30) and less negative (-.16) comments. It seems that those with the most negative comments made them because they least understood the reports; hence future professional development needs to concentrate on accuracy of reporting, as this would reduce the negative and highlight the positive aspects. A further analysis below will tease out more correlates of this interpretation score.

Differences between Teachers and Principals, and professional development or not professional development

Across all items there were more differences relating to whether participants did or did not attend professional development (Figure 9). Attending professional development increased familiarity with asTTle, enhanced the spread of information to other persons (students, senior management, board, and parents), gave more confidence in creating reading tests and entering scores, and increased the use of console, group learning pathways, tabular reports, and using the reports for future planning. Those who were not offered professional development claimed to prefer to learn at their own pace, teach themselves, internal rather than external professional development provision, and reading manuals. Clearly those who do not attend professional development find ways to defend their own self-learning.

Figure 9.
Effect-sizes for items that distinguished between those who did and did not receive Professional Development
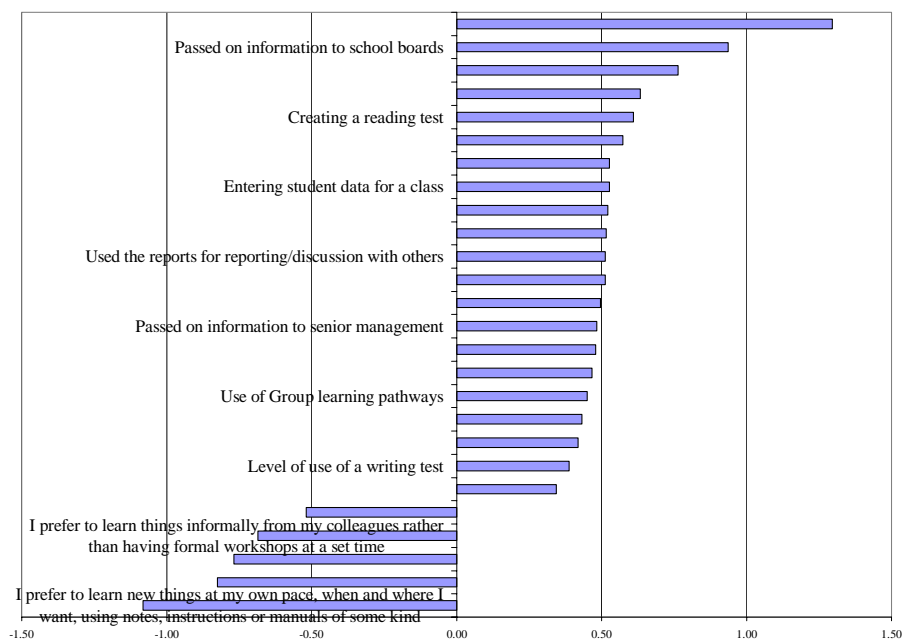
There were fewer differences between teachers and principals across all comparisons (Figure 10). Principals are more likely to pass on asTTle information to school boards, parents, senior management, syndicates, and students, and rated the Individual Learning Pathways and console report higher than teachers.

Figure 10.
Effect-sizes between principals and teachers
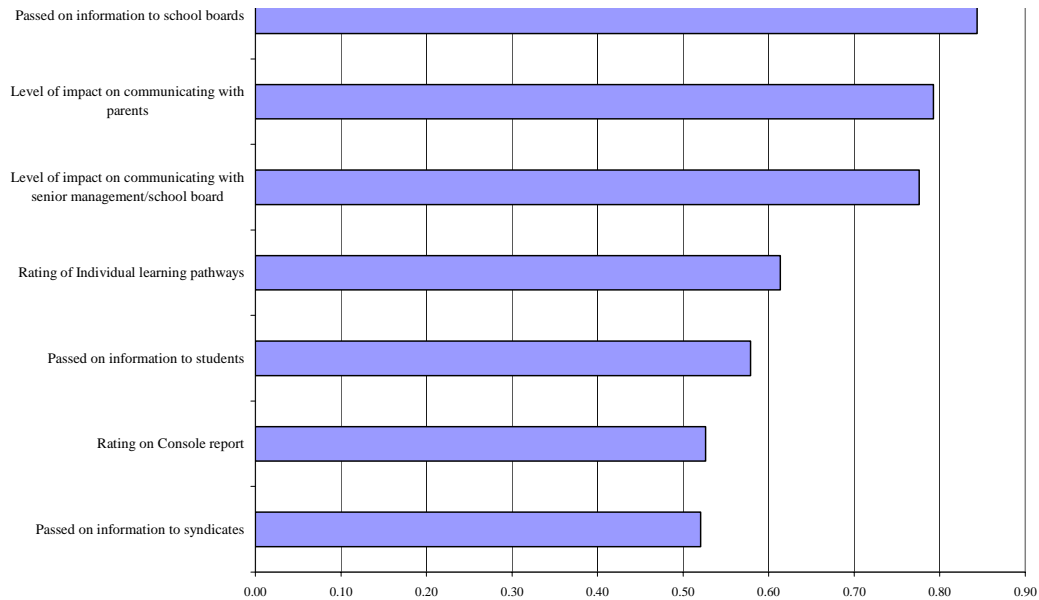
Relationships between the variables.

The 22 scales were divided into three major sub-sections, pertaining to precursors to using asTTle, the level of involvement with asTTle, and the value and worthwhileness of asTTle. Correlations of estimated scale scores within and across sections were obtained (Table 18). Statistically significant valuers are shown in bold, while missing values are shown as full stops. The greatest predictors of usage were confidence in using ICT and those who saw positive benefits from professional development. The most influential precursors to using and valuing asTTle relate to a positive attitude to ICT, the perceived adequacy of professional development received, and a conception that values assessment for student accountability. The greater the level and ease of use of asTTle then the higher the perceived benefits.

Table 18.
Relationships between the variables

		Precursors to using asTTle									Level and ease of use						Valuableness and accuracy						
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	Confidence in using ICT	1																					
2	Attitude to ICT	**.48**	1																				
3	Professional development: Prefer short training	.21	.	1																			
4	Professional development: Prefer on own	**.29**	.	**.29**	1																		
5	Professional development benefits	**-.28**	-.30	**-.46**	-.37	1																	
6	Professional development adequacy	.04	.00	**-.29**	**-.33**	-.05	1																
7	Assessment: for teaching	.18	.	.	.	-.07	-.12	1															
8	Assessment: school accountability	-.05	.	.	.	-.44	.11	.00	1														
9	Assessment: student accountability	-.23	.	.	.	.29	.03	.24	.24	1													
10	Ease of use	**.37**	.25	.03	.09	**-.49**	**.35**	**.31**	.00	-.02	1												
11	Total use of asTTle	**.22**	**.32**	.03	-.12	-.25	**.20**	-.01	**.36**	.05	.17	1											
12	Time Total'	-.05	-.09	-.06	-.26	.32	.05	.04	.32	-.10	-.15	**.31**	1										
13	Attitude: Easy to use	.04	.06	-.01	-.13	-.09	**.28**	.14	-.09	.21	**.39**	.14	-.13	1									
14	No. of positive comments	.12	.11	.09	**-.33**	-.18	.12	.18	.19	-.04	**.24**	**.39**	**.28**	.18	1								
15	No. of negative comments	.07	-.11	**-.26**	**-.41**	.22	-.03	.05	.19	-.22	-.17	**.25**	**.27**	**-.22**	**.49**	1							
16	Dissemination of information	.03	.23	-.05	**-.37**	-.10	.22	-.05	.13	.12	.04	**.45**	**.29**	.14	**.39**	.19	1						
17	Level of Impact	.04	.15	.10	-.16	**.38**	**.20**	.16	.20	.19	**.30**	**.22**	**.25**	**.27**	**.24**	-.07	**.30**	1					
18	Rating of six reports	.03	.23	-.04	-.21	-.22	**.39**	.18	**.39**	**.39**	**.48**	**.27**	.16	**.39**	**.36**	-.04	**.29**	**.55**	1				
19	Amount of new Information	.07	**.42**	-.12	-.23	.12	**.40**	.07	.10	**.32**	**.38**	**.23**	.03	**.35**	**.19**	-.13	**.35**	**.58**	**.71**	1			
20	Attitude: Useful	.05	.26	**.24**	-.12	.10	**.27**	.29	.17	.28	**.33**	**.23**	.12	**.59**	**.31**	**-.21**	**.19**	**.64**	**.58**	**.54**	1		
21	Attitude: Make a difference	.09	**.36**	**.29**	.08	.05	**.22**	.12	-.01	.24	**.34**	**.25**	.01	**.47**	**.28**	**-.27**	**.26**	**.66**	**.56**	**.63**	**.78**	1	
22	Interpretation score	.03	.14	-.01	**-.21**	.09	.13	**.34**	-.21	.20	**.19**	**.22**	.00	**.22**	**.30**	**-.16**	**.36**	.16	.21	.17	.17	**.18**	1

A maximum likelihood factor analysis with oblimin rotation was used on the level and ease of use, and valuableness and accuracy variables (the antecedents were not used given the missing values accrued because of the matrix sampling). There were four clear factors relating to (a) attitudes, (b) ease and confidence of usage, (c) level of usage and involvement, and (d) worthwhileness. The only variable that did not have clear loadings on any of the factors was the Interpretation– which at best loaded with level of usage —the greater the level of usage the more accurate the interpretations. Familiarity clearly breeds accuracy, and vice versa.

It is also worth noting that attitude was the greatest correlate of worthwhileness and value. Professional development sessions therefore need to be concerned with developing positive attitudes towards the innovation, and certainly to creating a sense of perceived usefulness. This is the same finding from previous evaluations of networked educational resources (Gilmore & Hattie, 1999 and Klobas, 1997).

Table 19.
Factor Structure

	Attitude	Ease	Usage	Worthwhile
Attitude: Useful	**1.02**	-0.11	0.10	-0.02
Attitude: Make a difference	**0.61**	-0.04	-0.02	0.34
Attitude: Easy to use	**0.57**	0.12	-0.03	-0.01
Ease of use	0.06	**0.94**	-0.15	0.17
Confidence in ICT	-0.02	**0.38**	0.07	-0.03
No. of positive comments	0.15	0.18	**0.76**	-0.04
No. of negative comments	-0.25	0.02	**0.73**	-0.13
Total use of asTTle	0.04	0.09	**0.43**	0.18
Time Total	0.06	0.14	**0. 35**	0.03
Interpretation accuracy	0.09	0.13	0.25	0.01
Amount of new Information	0.12	0.07	-0.11	**0.78**
Rating of six reports	0.23	0.20	0.07	**0.48**
Dissemination of information	-0.12	-0.05	0.39	**0.45**
Level of Impact	0.40	-0.01	0.05	**0.43**

Factor Correlation Matrix				
Attitude	1.00			
Ease	0.28	1.00		
Usage	0.08	0.10	1.00	
Worthwhileness	0.54	0.17	0.28	1.00

Causal relations between attitudes, usage and correctness of interpretation.

Structural equation models allow for the assessment of various models predicting the level of use and value of asTTle. A structural equation model was specified between three levels of variables. The antecedents included various attitudes that could impact on the intention to use asTTle: (all variables have been introduced above)

Confidence using a computer;
Attitude to ICT as part of teaching and learning;
Preference for PD courses;
Preference to do PD on their own;
Concept of assessment – for teaching;
Concept of assessment – for school accountability;
Concept of assessment – for student accountability.

The second group of variables included those that were more directly related to level of use of asTTle:

Ease of Use (measured via two different questions);
Total use of asTTle;
Total time spent using asTTle.

The third set related to the perceived worthwhileness or value of asTTle, with particular emphasis on the use and level of impact of the reports.

Dissemination of information;
Level of impact;
Rating of six reports;
Amount of new information;
Attitude to use of asTTle;
asTTle will make a difference.

Finally, the interpretation score was included, and for each set, a path was included to assess the influence. The final model provided good fit overall, but there were some variables that did not contribute to the model. For example, the conception of assessment relating to school and student accountability did not contribute as antecedents to the use or otherwise of asTTle. The total time spent using asTTle was not a powerful predictor compared to the other usage factors. The model was re-run omitting these variables and the final solution provided in Figure 11. The goodness of fit of this model (chi-square =521.09, *df*=54) can be represented by the Normalised Fit Index of .92, and Tucker-Lewis index of .89 – which is most acceptable given this is a path and not structural model (see Bollen, 1989).

The most powerful antecedents related to the teachers' attitudes towards using ICT as part of teaching and learning. Other influences on the antecedents were the generalised attitude to computers, a conception the assessment can be powerful to influence teaching, and preference for short or self professional development. Of no effect were conceptions of assessment related to school or student accountability. The three variables relating to ease of use and total time contributed similarly to level of involvement with asTTle. Five of the six variables contributed most to the "welcoming" of asTTle, and the one with the lower contribution was whether the teachers or principals actually disseminated the reports. Overall, those teachers who welcomed asTTle as something that "made a difference" to them had higher accuracy of interpretation of the asTTle reports.

Figure 11.
Structural Model
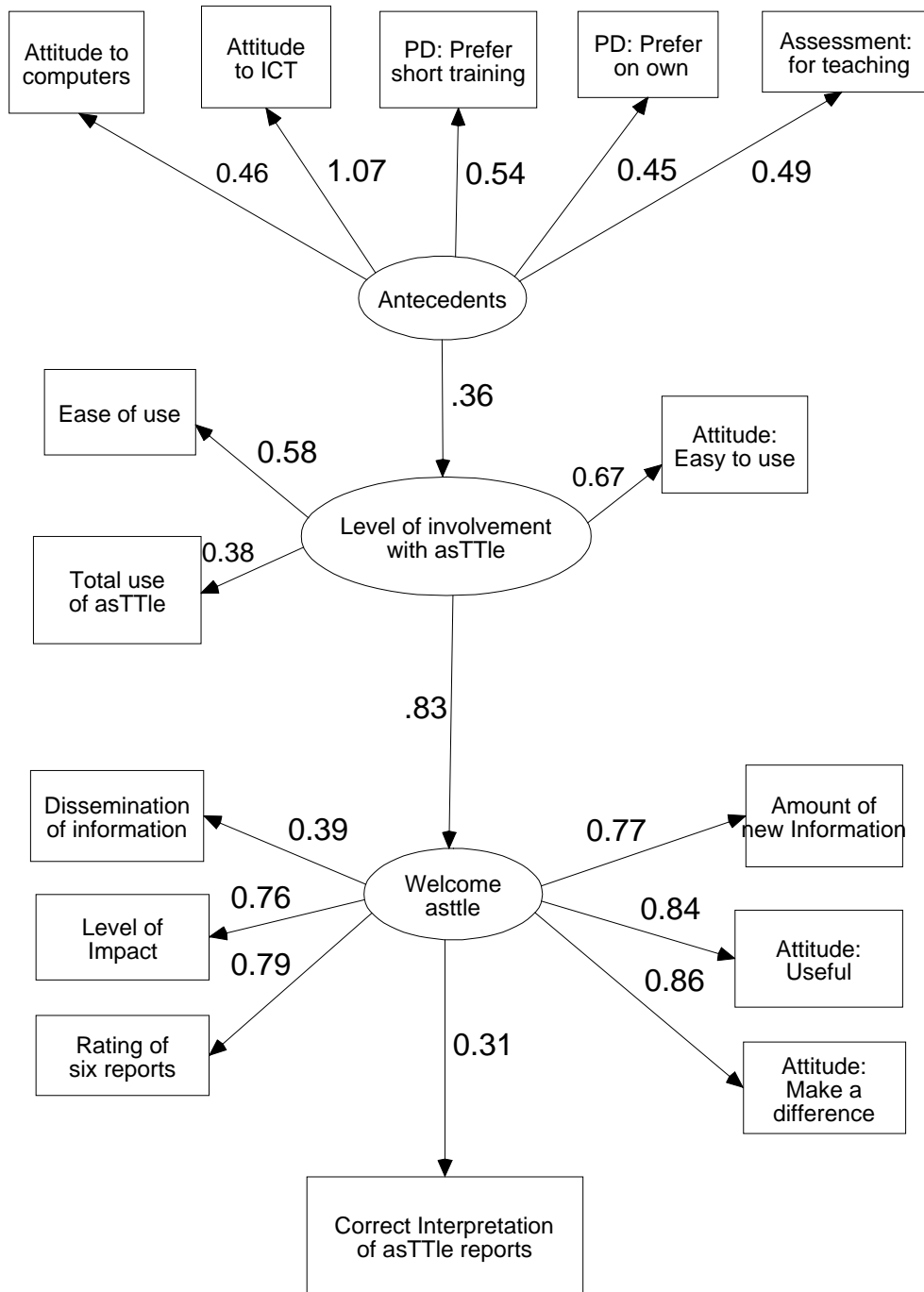
The major messages from this structural model are that professional development needs to be oriented most towards encouraging an attitude of the value of ICT in teaching and learning. By positively influencing the teachers' attitudes towards ICT

there is a higher likelihood of correctly using asTTle as part of teaching and learning. It is noted, that many of these New Zealand teachers came to the asTTle application with already high levels of such attitudes to ICT as important in teaching and assessment. Further, the accuracy of reports is more a function of the level of impact, perceived quality of the reports, and belief that the results from asTTle will make a difference. This underlines the importance, as outlined in the Theory of Planned Behaviour, of perceived usefulness to subsequent use and correctness of use.

Professional development courses (whether run externally, internally, or self-presented) needs to attend to perceived usefulness of innovations such as asTTle. They need to attend to the positive attitudes towards ICT or these will become a block to the use of asTTle, and then ensure that the teachers find the application easy to use. The major attention needs to be devoted to the usefulness, new information, ways the reports can have impact, and how asTTle can make a difference to teaching and learning. Priority needs to be given to the correct interpretation of the reports. From the professional development offered as part of this trial, it appears that all but the correctness of the reports was successfully covered.

Conclusions

asTTle is a different way of considering assessment, both at the national and at the school and classroom level. It provides considerable control to the user in determining how it is used, and more importantly, how it is interpreted. As it is aimed at influencing teaching and learning rather than the accountability of teachers and schools, it is different to the usual national testing models in other countries. As predicted by the Theory of Reasoned Action, the attitudes and intentions of the users, the subjective norms (such as principals) that these people interact with, and the self-confidence in using and interpreting the tool were the critical predictors of the level of use and accuracy of use of asTTle.

In these trials, there was high adoption of the application – most teachers complied with the request to create, administer, and interpret at least a Reading and Writing test, and most shared the results of these assessments with others in the school and beyond. Certainly, asTTle was easy to use.

The teachers preferred whatever professional development they received (either by a professional development team, by school-based initiatives, or by self-study of the resources within the asTTle tool). Those who had professional development (from AToL) considered the primary benefit related to interpreting the reports, and those not attending suggested that the greatest perceived benefit if they had had professional development related to using the interpretations from the reports. This highlights the importance of any professional development (particularly externally provided professional development) ensuring a high level of correct interpretation of the reports. Similarly, the level of correct interpretation was the major difference between those who did, or did not, receive professional development.

The major interpretation errors related to the relationships between the scores and the quadrants in the Individual Learning Pathways. The mistake most often made was to assume that the Strengths or Gaps (etc.) listed were relative to the national New Zealand sample. That is, the interpretation was that a particular concept (e.g.,

Understand Main Ideas) was listed as a strength because relative to the New Zealand national sample this student could easily meet this objective. Instead, the correct interpretation is that this student could get items relating to Understanding Main Ideas correct, whereas it would have been expected that these items were actually more difficult than the overall ability of THIS student – hence they are a strength relative to the overall level of ability of the particular student. There is NO nationally normative data used in the creation or interpretation of each of the four quadrants in the Individual Learning Pathway.

We consider that the Group Learning Pathway could have contributed to this interpretation error, so completely redesigned this report. We trialled the new report on teachers and they were less prone to making the above error of interpreting the report compared to some overall group or national norm. Certainly, professional development needs to attend to the correct interpretation of this report. It may be that self-administered professional development packages relating to the correct interpretation of the reports could be devised and made available on the asTTle (TKI) web site, and then be used by the various professional development external providers around the country.

A major set of findings from this evaluation relates to the effects of professional development. All too often, the claim is that professional development is effective and needed, but so many teachers found ways to "teach" themselves using the resources provided within the tool and from their immediate colleagues. The effects of professional development are more specific to certain content and skills, and it is therefore important that future professional development becomes oriented more to content and skills that have been shown to be important outcomes from such external professional development. In the current case, professional development was less related to adoption of the innovation but to specific ways it was used and whether or not it was used correctly. Professional development had positive effects on the level and ease of use of asTTle, on the degree of presenting the asTTle reports to others (e.g., parents, principals, school boards), and on the value of each of the reports. Possibly the most important difference was that those who attended professional development were more likely to use asTTle when talking with students, in their planning for the class and for each student, and were more likely to correctly interpret these reports.

The respondents claimed that asTTle would have major effects on their teaching and planning, on the way they assessed their students, and on communicating with students, management, and parents. They noted the power and value of new information now available on performance relative to national norms, student leaning needs (strengths, weakness, etc.), and appropriate teaching tools (the What Next web site is the most "hit" website of all). The strengths of asTTle were the quality and type of information, the ease of use, the flexibility of the tool, and the content of the tests. The weaknesses mostly related to time needed, the costs of photocopying the tests, data transfer, and some technical concerns. These are being addressed in future versions, as far as is possible. For example, the specifications for importing data from the various data management systems used in New Zealand schools (there are 14 approved systems) have been sent to the suppliers, with some already incorporating an "export from the management system" that can then be imported into asTTle – thus reducing the time to input the student information. The problems of

photocopying still remain, but effective planning can reduce one of the problems of the trial. That is, the teachers were expected to photocopy the tests using their already allocated monies for printing (recall the trial was in September), but school principals may need to plan allocating school-based resources for printing separate from individual teachers' allocations (and use both sides of the page thereby halving paper costs). There is a reluctance to provide answer sheets separate from the test booklets (as requested by some in the trial) as this would lead to cupboards of previously designed asTTle tests (as noted by many of the same respondents). This would defeat a major advantage of asTTle – that it can be tailored to the teaching that has just occurred. The problems of scoring will be ameliorated as asTTle moves to a web based student administered version.

From this evaluation many improvements can be made to the next version of asTTle. These involve (a) more optimal spacing, (b) data sharing, (c) improved writing and overall manuals, (d) improved linear programming heuristic that controls the choice of items, (e) communicating data with student management systems and (f) inclusion of an improved group learning pathways report.

As with most innovations in schools there will be a time of initial capture, particularly given that the assessment tool is in the "control" of the school-based users, it is informative to them, and it is seen as a valuable addition to their teaching repertoire. Ultimately, however, the continued use will relate more to the usefulness and power of asTTle reporting, to the provision of information that could not more easily be accessed from other sources, and to the quality and dependability of the psychometrics underlying the tool. As predicted by the Theory of Reasoned Action, the perceived usefulness, the subjective norms from important others (e.g., Principals, Education of Review Office, Ministry of Education), and self-efficacy towards using computers as part of teaching make the difference to the impact of innovations such as asTTle.

References

Azjen, I. and Madden, T.J. (1986). Predictions of goal-direct behavior: Attitudes, intentions, and perceived behavioural control. *Journal of Experimental Social Psychology, 22,* 53-474.

Brown, G.T.L. (2002). *Teachers Conceptions of Assessment,* Unpublished doctorial dissertation, University of Auckland, Auckland, New Zealand.

Bollen, K. (1989). *Structural Equations with Latent Variables.* New York: John Wiley & Sons, Inc.

Fishbein, M. & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.

Gilmore, A. & Hattie, J.A. (2001). Understanding usage of an internet based information resource for teachers: The Assessment Resource Banks. *New Zealand Journal of Educational Studies, 32*, (2), 237-258.

Hambleton, R.K., Swaminathan, H. & Rogers J. (1992). *Fundamentals of item response theory*. Newbury Park, CA: Sage.

Klobas, J. (1997). *A behavioural intention model of networked information resource use*. Unpublished doctoral dissertation, University of Western Australia, Perth.

Lavery, L. & Brown G.T. (2002). *Overall summary of teacher feedback from the calibrations and trials of the asTTle reading, writing, and mathematics assessments 2002*. asTTle Technical Report #33. Auckland, New Zealand: University of Auckland, project asTTle.

Tyack, D. & Cuban, L. (1995). *Tinkering toward utopia: A century of public school reform*. Cambridge, MA: Harvard University Press.