Reporting Assessment Information to Teachers: Report of Project asTTle Outputs Design

Technical Report 15

Gavin T. L. Brown
University of Auckland

This report was prepared to aid the design of Project asTTle outputs for reporting assessment information to teachers. Issues in graphical communication, user-interface preferences, and communication of statistical information to parents, administrators, and teachers were reviewed. The report overviews the three major output reports that will communicate student achievement in literacy and numeracy in the forthcoming Project asTTle CD-ROM and identifies how the issues raised are addressed.

Table of Contents

Project asTTle (Assessment Tools for Teaching and Learning) offers teachers a software tool to assess the achievement of students in Years 5 to 7 against the criteria of the New Zealand numeracy and literacy curricula, and against norms for a nationally representative sample. The software will analyse test scores and provide automatic reports for teachers on the strengths and weaknesses of individual students and how well those students performed in relation to others. This report examines the literature on the level of teacher and parent understanding of assessment results, their preferences for types of reporting mechanisms, and user-interface requirements for communicating information in both on-line and hardcopy media. Recommendations from the literature and guidelines developed from user evaluations were used to design the asTTle output reports.

Literature Review

Reports and Grades

Understanding and communicating assessment results has been identified as one of the most critical teacher professional development needs in the field of educational measurement expertise (Plake, Impara, & Fager, 1993). Grade symbols, which are perhaps the most common means of reporting assessments, are complex communications (Waltman & Frisbie, 1994). They contain three facets: (a) reference, (b) status, and (c) content. The first facet refers to whether the point of comparison or reference that the grade refers to is either a population norm or a criterion or construct standard. The second facet, status, indicates that a grade refers either to the present status of a student's achievement or to progress from some time in the past. The last facet focuses on whether the content of a grade refers to purely academic achievement or is blended with non-academic characteristics (such as effort, behaviour, attitude, or appearance) as well. Stiggins, Frisbie, and Griswold (1989, p. 6) argued against reporting growth because progress is "very difficult to measure and evaluate accurately and equitably". Likewise, they argued against the use of student aptitude, interest, personality, learning ability, attitude, or effort as a factor in reporting achievement or resolving borderline cases. According to Guskey (1996), the assigning of letter grades is too subjective and error-prone. He argued instead for the practice of relating grades and reports to learning criteria (i.e., product,

process, and progress) so that a clear picture of learning is generated. Reports should indicate what students know (product), how students learned or demonstrated it (process), and how much they have learned (progress).

In light of this, Waltman and Frisbie (1994) recommended that grades report one thing only from each facet pair so that grades report present status of academic achievement only, and that they indicate clearly whether they are criterion or norm referenced. They recommended reporting separately academic and non-academic achievement. Friedman and Frisbie (1995) recommended that separate subheadings of traits be reported. This would allow diagnostic information to be communicated to parents so that they would better understand the nature of a child's performance in a broader content domain. Guskey (1996) recommended reporting of product criteria because these criteria are much more objective and reliable, though clear indicators for process and progress may make reporting those aspects of student achievement viable.

Guskey (1996) argued that grades are beneficial only if they are supported by specific or individualised feedback from teachers. He drew attention to the tension between grading and checking, asserting that grading is not essential to instruction, whereas checking regularly on learning progress is. This is so because checking is diagnostic and prescriptive, while grading is descriptive and evaluative. Guskey (1996) further drew attention to the fact that no one method of reporting suits all audiences (i.e., parents, teachers, students, administrators) nor can one grading method meet all purposes of a report. Although brief letter grades are quick to create they do lack information about the nature of student achievement. On the other hand, narrative reports, despite their rich information, are extremely time-consuming to prepare and assimilate. Guskey (1996) concluded with three main guidelines about reporting: (a) good reporting depends on having a clear purpose, audience, and consequence; (b) with the intended audience in mind, accurate and understandable descriptions of student learning are needed; and (c) reporting methods need to enable teaching and learning (this means not averaging grades across learning areas or assigning zero for late, missing, or neglected work).

Lake and Kafka (1996, p. 90) emphasised that "recording of information about a student's progress and achievements should serve a useful purpose and, in particular, should benefit the student". For them, all reporting needs to be developmentally appropriate. This means that checklists have to be sensitive to the learning expectations appropriate to the age and stage of students being reported. Descriptors of achievement levels should give a picture of the developmental sequence so that teaching and student work is guided. Although the rationale is different from that of other authors, this recommendation is not inconsistent with recommendations on describing student product achievement or profiles.

Visual and Text

Jaeger, Gorney, Johnson, Putnam, and Williamson (n.d.) and Jaeger and Putnam (1994) reported results of a survey of parents, school board members, media representatives, school superintendents, and district information officers about school report cards, not individual student reports, that provided important insights. Contrasting prototype printed reports were trialled in a two-by-two (length-by-format) framework: long (4-page) vs. short (1½-page) reports and tabular/graphic vs. narrative reports. Respondents were asked to indicate preferences for format and length as well as content. In addition, comprehension questions were posed to determine which report type provided the most accurate communication of information. In terms of preference, the long tabular/graphic reports were superior, but in terms of accuracy, the long narrative was superior for all respondents. Jaeger and Putnam (1994, p. 25) concluded that "a design that combined the judicious use of tabular and graphical data displays with redundant narrative discussion and rich interpretation would appear to be even better."

Wainer, Hambleton, and Meara (1999) redesigned graphic data displays from the 1994

National Assessment of Educational Progress (NAEP) reading assessment to improve their communicability. In a controlled experiment, it was found that visual displays redesigned according to Wainer's (1997) principles yielded more accurate and faster answers. The principles advocated include: (a) displays should allow access to the user's deep structure without recourse to linguistic decoding – that is, they should be "seen" and not "read"; (b) displays should have a clear communicative purpose; (c) displays should not attempt to do too much; and (d) displays should avoid clutter (e.g., footnotes, notations). Respondents indicated a preference for bars rather than lines and for the use of colour. In addition, visual displays that featured "bigger print, additional white space, [and a] "lack of … footnote[s]" (Wainer et al., 1999, p. 319) were also preferred. Accumulated displays and footnotes were seen as inappropriate despite the fact that they may display information in a more accurate way. Finally, it was concluded that conventional displays with which users were most familiar tended to be preferred.

In a study with a similar respondent profile, Henry (1993, p. 76), found that graphic displays were "a reasonable option for information-rich displays" in that the level of understanding, though poorer than table-based information, was still reasonable. He reasoned that practice with graphical forms would enhance ease of understanding and argued that comparison information must be in close proximity to ensure accuracy of comprehension. Another important graphical principle supported by Henry's findings was the importance of making sure empty space is not used to communicate information: "gaps do not necessarily convey quantities accurately to the viewer" (1993, p. 76).

An interesting application of these graphical reporting principles is the Quality School Portfolio (QSP) (Baker, 1999). The QSP uses computer technology to communicate school quality information using a series of dynamic dials and gauges that show the relationship of the school to district norms. This allows administrators and parents to get a quick overview of key information about a range of schools.

Teachers' Interpretation of Assessment

Impara, Divine, Bruce, Liverman, and Gay (1991) evaluated the ability of teachers to interpret a hypothetical student's test score information by providing reports that either had or did not have interpretive information. The availability of interpretive information and previous participation in a measurement class were significant predictors of ability to accurately interpret score results. For most teachers, the ability to interpret the percentile band performance profile was low.

Linn and Dunbar (1992) showed that the anchor item exemplars of levels of achievement from the NAEP are easily misunderstood. Hambleton and Slater's (1997) research on NAEP executive summary reports further highlights the problems policy makers and educators have in understanding reports of assessment. Despite the fact that three-quarters of the sample had one or more statistics or testing courses in their background, "many interviewees had forgotten a lot of the statistical and measurement information they had known at one time" (Hambleton & Slater, 1997, p. 8). Problems were found with the statistical jargon or terminology used (e.g., "proficient", "statistically significant", "standard error", "cutpoint", "scale score"), the construction of tables (e.g., cumulative columns exceeding 100%, too much detail, statistical symbols like greater than and less than signs, ordering of material, footnotes), and the design of graphs (e.g., over-complexity, novelty). Guidelines prepared by Hambleton and Slater (1997) suggested that all graphs, figures, and tables be: (a) understandable without reference to text; (b) simple and straightforward; (c) field-tested with focus groups representing the intended audience; (d) drawn so that quality remains consistent even when reproduced or reduced; and (e) designed specifically for the intended audience. The authors concluded with a plea for rigour in the design of the user-interface for assessment reports: "Ways need to be found to balance statistical rigor and accuracy in reporting with the informational needs, time

constraints, and quantitative literacy of intended audiences" (Hambleton & Slater, 1997, p. 30).

Salvagno and Teglasi (1987) found that teachers preferred interpretive rather than factual information in psychological reports about students. Interpretive information contains analysis, synthesis, or the implications of test results, while factual information is quantitative or consists of short statements.

Reporting to Parents

Barber and Evans (1992) reported that few states or districts require reporting of test results to parents. They found that where reporting to parents was carried out it usually consisted of an individual report to the home that included only the percentile rank, the percentage correct or the percentage of objectives passed. Some reporting to parents was conducted through conferences or newsletters. In order to make accurate interpretations of results, parents need information about progress since last testing, strategies they can use to help students perform better, and narrative explanations of the relationship between test results and classroom practice and achievement. Many parents misunderstand scores, interpreting criterion-referenced results (especially if values are high) as norm referenced information.

Waltman and Frisbie (1994) found that teachers and parents did not agree as to what a reported grade meant and interpreted the meanings of each grade differently. For example, teachers tended to assign "C" as the lowest grade, while parents interpreted this as an average or acceptable grade.

Checklists of learning outcomes are overwhelming to parents. Checklists do not necessarily indicate whether the student result is in line with the expectations for that level (Guskey, 1996).

Jaeger and Putnam (1994) reported that parents preferred long reports over short ones, with a slight preference for tabular formats They recommended, when reporting about schools to parents, that rich and varied indicators of quality be provided, that a mixture of narrative, graphical, and tabular formats be used, and that redundancy of information in different formats is warranted.

Summary

The communication of assessment information to teachers, administrators, and parents is a difficult task that can be aided by judicious use of graphical communication principles. These principles have to do with designing reporting mechanisms in accordance with the kind of information, the purpose of the information, and the audience being addressed. The judicious testing of designs is required (i.e., testing both preference and accuracy). Information displayed in graphic form should be supplemented with appropriate and concise interpretive text that will engage and assist the audience.

Design Procedures

Consistent with recommendations from Wainer et al. (1999), pilot testing of visual displays was carried out with intended users. Focus groups were used to discuss issues pertaining to preferences for the design of the visual displays and the nature of intended users' understanding. Procedures for testing the reporting design followed Scriven's (1991) procedure of internal alpha testing, including project personnel and Ministry of Education personnel, and external beta testing. Additional constraints on the design of the output existed in the form of programming requirements and budgets that reduced the desired levels of dynamic interaction.

The major design requirement of Project asTTle reporting was to provide teachers with high-quality, externally referenced information about the achievement of students, as well as detailed, reliable information about the nature of student achievement. Thus, reporting outputs had to fulfil both normative and criterion referenced interpretations. Secondly, the outputs would initially be encountered in a graphic online environment that teachers could explore before determining which information would be printed or transferred to school databases that are used for management purposes. This meant that an interface with some dynamic features and with high-quality

user-interface characteristics (Spolsky, 2001) had to be developed.

Research was carried out to determine the nature of comparisons teachers would like to be able to make (Meagher-Lundberg, 2000), and to determine user reactions to various drafts of the interface formats (Meagher-Lundberg, 2000, 2001a, 2001b). Those reports guided the development of the interface content and design. The types of comparisons that teachers like to make were included in the comparison console selection options. Teachers had problems in understanding how to navigate from one display to another; these problems were resolved by placing interfaces in a web browser display. Teacher difficulties in understanding output content and labelling were identified and resolved.

Interfaces for Reporting Achievement

Several output templates were designed: (a) an achievement comparison console; (b) individual and group "kid maps"; (c) a curriculum level achievement "exploded pie charts"; (d) across-time comparison bar graphs; (e) a cognitive processing achievement bar graph; and (f) an attitude comparison bar graph. It should be noted that the achievement comparison console contains a number of the reports including the inter-time comparisons, cognitive processing achievement, and attitude comparisons. The characteristics of each report template are discussed below.

Data are entered into the asTTle software after student tests have been scored according to marking keys provided by the package. Once analysis of the scores is completed, the user is presented with an interface that displays the achievement of the students within a standard web browser interface, complete with drop-down menus, iconic tool bars, and standard web navigation tools and menus (Figure 1). The main console automatically reports student performance on curriculum areas previously selected by the teacher in the design of the test, attitudes, performance by cognitive processing requirements of questions, and achievement by Year level relative to the national norms, previously obtained through standardisation.

The main data presented in Figure 1 are the achievement of students on various curriculum categories using a *standard-900 scale* (where the mean = 500 and the standard deviation = 100). The data are presented in a calibrated and coloured semi-circular dial where values 100 to 400 are coloured red, values 400 to 600 are yellow, and those above 600 are green. The colour scheme was adopted, despite initial design reluctance, because of strong user feedback (Meagher-Lundberg, 2001a). The pointer indicates the estimated achievement of students relative to the national norm.

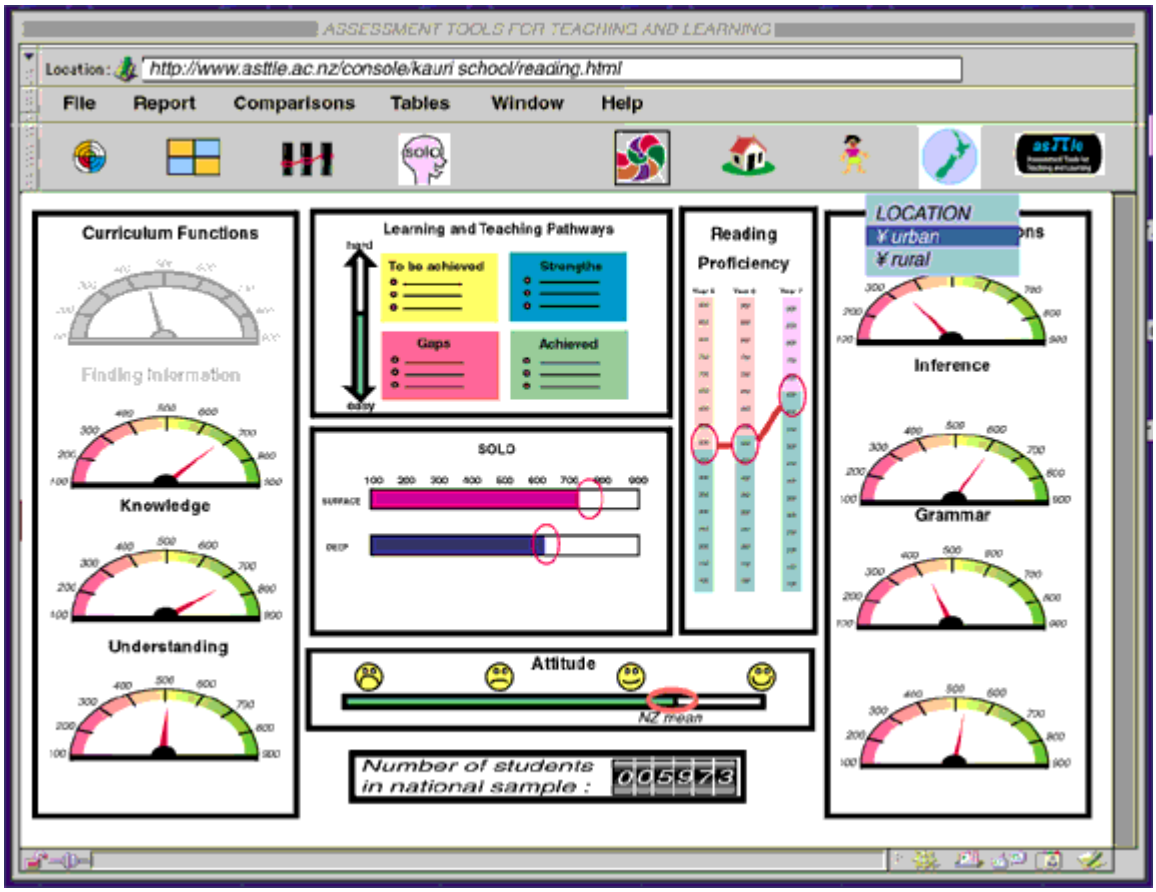

Figure 1. Project asTTle main console showing hypothetical school results compared to urban schools.

The attitude of students towards the curriculum areas of reading, writing, and mathematics is measured through a set of six questions adopted from the National Education Monitoring Project (NEMP) survey questions. Students indicate their attitude by selecting one of four "happy" faces, as shown in Figure 1. A red circle against the horizontal green bar represents the average score of students against the national mean.

The achievement of students across differing Year groups is reported against vertical bars, which are shaded green up to the national mean for each Year level. Scores are reported against the standard-900 scale score used in the semi-circular dials. In the event that all students being reported are from the same Year group, then only one bar would be displayed. Provision is made for up to three Year groups to be reported at one time. Red circles joined with a red line indicate results for the group being reported. The position of the circles relative to the Year mean allows quick identification of whether one Year group or another is further from or closer to the national mean.

In order to account for the standard error of measurement, and to reflect Hambleton and Slater's (1997) findings regarding the misinterpretation of technical information, it was decided to indicate achievement with a red circle or oval. The size of the measurement error is indicated by the size of the circle; thus, the smaller the circle, the more accurate the value of the reported achievement.

Students' cognitive processing is indicated by the percentage of items, as classified according to the SOLO taxonomy (see Meagher-Lundberg & Brown, 2001), answered correctly. It was decided to summarise achievement on the SOLO taxonomy by reporting achievement against the New Zealand norm for surface (i.e., unistructural and multistructural) items and deep (i.e., relational and extended abstract) items. Part of the reason for this aggregation is the difficulty teachers have in comprehending the SOLO taxonomy (Meagher-Lundberg, 2001b), and the relative

paucity of items written at the extended abstract level (Meagher-Lundberg & Brown, 2001). Each pair of bars shows the national mean and the group achievement by the four SOLO categories. As in the attitude bar, a red circle against the horizontal green bar represents the average score of students against the national mean. Relative performance is indicated by the standard-900 scale score used in the semi-circular dials report.

All the report formats discussed above are used to indicate student performance relative to that of the norm, whether that is the national norm or a sub-population norm. However, as reported by Meagher-Lundberg (2000), teachers clearly indicated a desire for some sort of "absolute" interpretation of what students actually know, understand, or can do. Furthermore, they preferred this information to be related to the achievement objectives and levels of the pertinent curriculum statement. To achieve this end, several other report formats were designed.

The key curriculum achievement report (Figure 2) is the *kidmap* that shows both an absolute and a relative description of a student's achievement. The relative dimension is shown in the red circle indicating the achievement of the child on the standard-900 scale that also has the New Zealand norm clearly marked. The absolute achievement in terms of curriculum categories is displayed through a four-quadrant map, wherein the upper and lower halves represent hard items and easy items respectively, while the left and right halves represent the items answered correctly and incorrectly respectively. Through item response theory (single parameter) analysis it is possible to locate the student's estimated achievement relative to the difficulty of the calibrated items. The mid-space between the upper and lower halves represents the estimate of the child's ability, also signalled by the red circle on the standard-900 scale.

The quadrants are clearly marked both with labels and colours that indicate the nature of the items present in each quadrant. Each quadrant contains the appropriate curriculum descriptors, previously obtained from the item signature study (Meagher-Lundberg & Brown, 2001), and shows the item numbers associated with each descriptor.

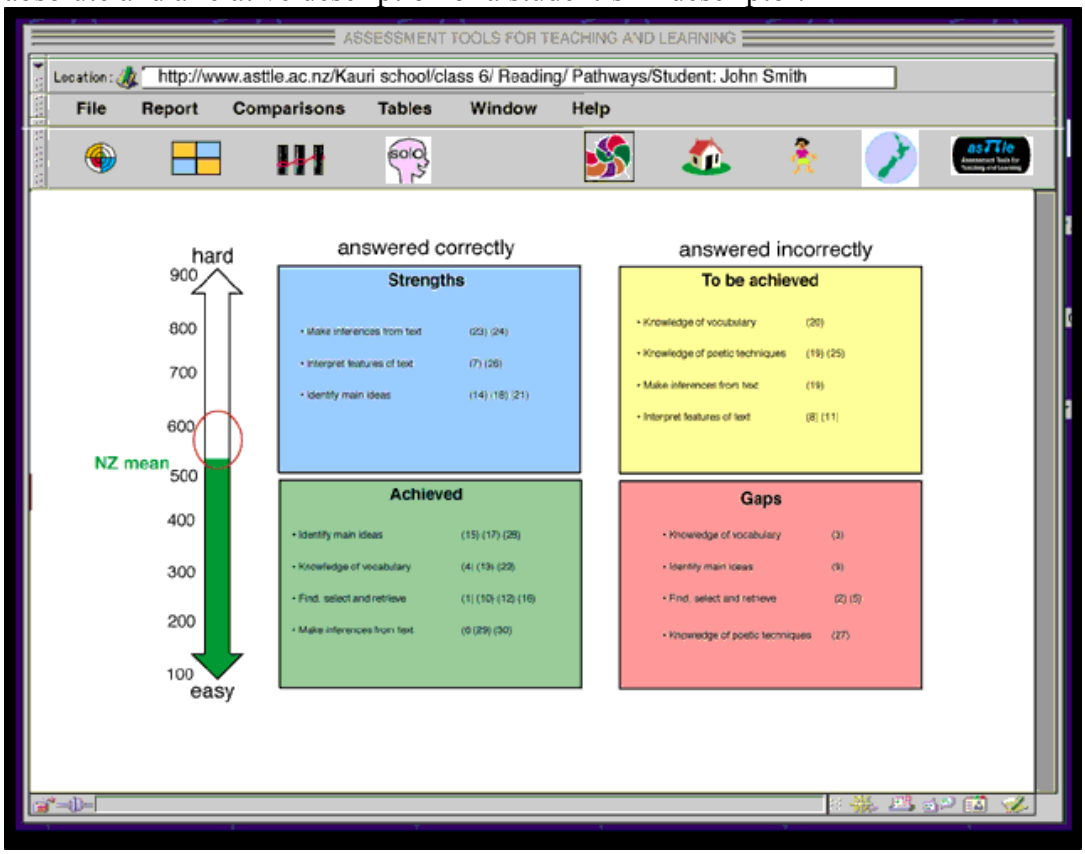

Figure 2. Project asTTle *kidmap* showing hypothetical student results for reading assessment.

The light-blue, top-left quadrant, marked "Strength", shows items that the student answered correctly but which were more difficult than the estimate of the student's ability. This quadrant thus displays the child's unexpected strengths that should be exploited in future teaching and learning. The light-blue colour was chosen to give a strong "the sky's the limit" signal to the teacher.

In contrast, the yellow, top-right quadrant, marked "To Be Achieved", shows the curriculum descriptors and items that the student answered incorrectly and which were more difficult than the estimate of the student's ability. These items represent areas that the child still has to achieve and in which it is expected the teacher will carry out more teaching. The yellow colour reinforces the need for caution and more instruction.

The light-green, bottom-left quadrant, marked "Achieved", displays the items and curriculum descriptors that were relatively easy for the student in relation to the estimate of their ability and which were answered correctly. The green colour, suggestive of a green traffic light, signifies that these are areas in which the teacher can confidently give the student more work and which should no longer dominate instructional content or time.

In contrast, the red, bottom-right quadrant, marked "Gaps", displays items and curriculum descriptors that were relatively easy in relation to the estimate of the student's ability but which were answered incorrectly. The colour red clearly signals that this is a danger area that the teacher needs to investigate to determine the nature of the gap in the student's achievement. Possible explanations, as alternatives to the student not actually knowing the material, include such things as carelessness, skipping items, illness, and so on. Nevertheless, these are items that the teacher ought to investigate and either eliminate as a concern or address in a remediation plan.

Figure 3 shows a second type of absolute information report – this time achievement reported by curriculum strand and level rather than by asTTle scale score. Earlier studies (Limbrick, Keenan, & Girven, 2000; Glasswell, Parr, & Aiken, 2001; and Ell, 2001) provided analyses of the literacy and numeracy curriculum documents that were used in the item signature reports (Meagher-Lundberg & Brown, 2001a, 2001b), to map the curriculum content of each assessment item. Other studies (Hattie, Tognolini, Keegan, & Brown, in progress) mapped all items to the relevant curriculum level and stage within each level. With this information, it is possible to display achievement to curriculum level by curriculum content area. The display uses an exploded pie chart to separate curriculum content areas, and concentric circles to indicate curriculum level of achievement. The percentage of students in the assessed group at each curriculum level is indicated by a value in the appropriate cell for each assessed curriculum area. Blank areas indicate that the teacher chose no assessment material for that area. The teacher can obtain a list of students by level and curriculum area by clicking on the curriculum area label.

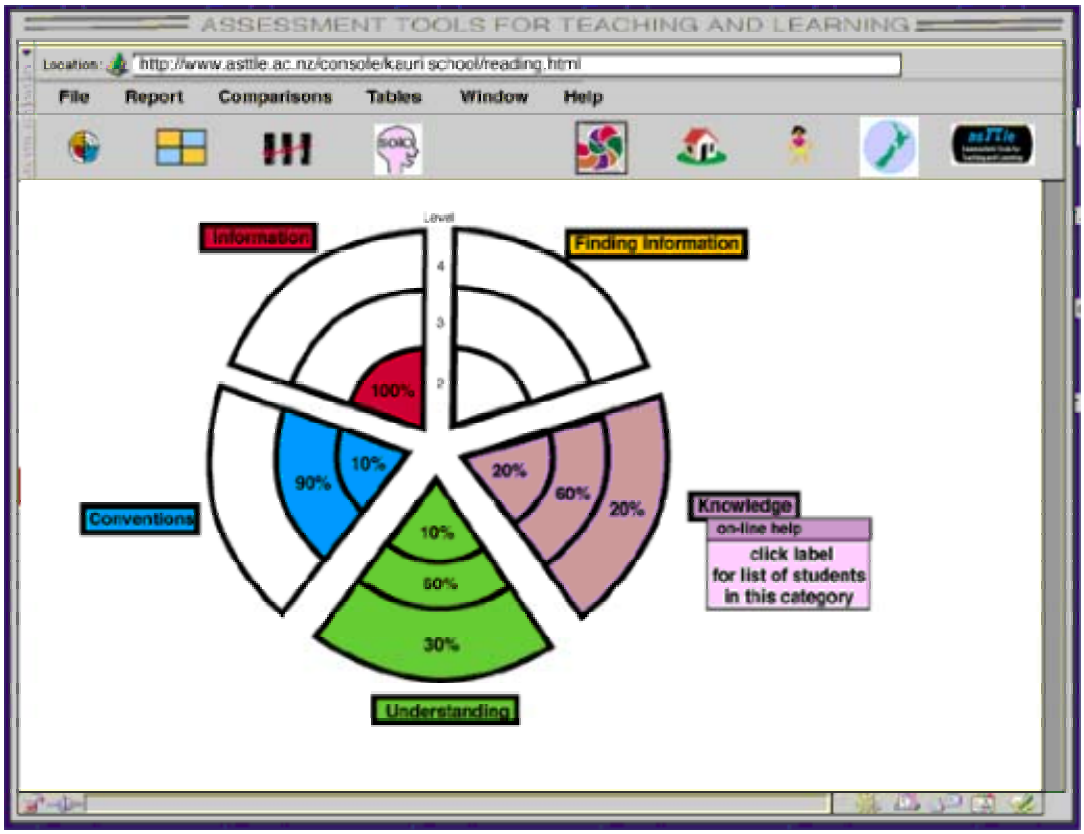

Figure 3. Project asTTle "exploded pie chart" showing hypothetical school results by curriculum levels.

Conclusion

Preliminary investigations on the communicability of the asTTle report interfaces (Meagher-Lundberg 2001a, 2001b) suggested that teachers would understand the information reported. However, further work is planned in conjunction with the field trialling of the prototype to ensure that not only will teachers understand how to use the software in which the reports sit but also have an accurate grasp of the information reported.

Nevertheless, this study shows that it is possible to design interactive reporting formats that meet theoretical, psychometric, and user-interface requirements. Specifically, these reports indicate clearly the type of information being reported. That is, it is clear (a) whether a norm or criterion reference is being made; (b) that curriculum achievement, and not student attitude or effort, is being reported; (c) that the architecture or traits of achievement are being reported; and (d) when progress over time, rather than a time-marked snapshot, is being reported. Multiple report formats allow communication of information to a variety of audiences, including principals and parents. This information, in conjunction with the curriculum map and item characteristics information, allows teachers to identify potential learning needs and work towards improved student understanding and ability.

The highly graphic nature of asTTle reports permits viewers to access the deep structure without having to engage in a great deal of linguistic decoding. Because multiple reports are possible, each display attempts to communicate a limited range of relevant information. Through appropriate use of online help files and popup balloons, the amount of clutter (e.g., footnotes, notations) needed to explain the graphs has been reduced. The use of bars, colours, and empty space should also permit the user to accurately and rapidly grasp the information presented. The use of conventional display frameworks (i.e., the browser) and conventional colour schemes (e.g., red for low or below expected) has enabled more rapid communication of information.

The use of connected circles, overlaid bars, and circles of various sizes to communicate various statistical concepts also seems in accordance with the research. The statistical indicators of change, means, and standard error measurements are calculated behind the scenes and communicated graphically. This is an appropriate level of sophistication given the informational needs, time constraints, and quantitative literacy of New Zealand teachers.

References

Baker, E. L. (1999). *Technology: Something's coming – Something good* (CRESST Policy Brief 2). Los Angeles: UCLA Graduate School of Education and Information Studies, National Center for Research on Evaluation, Standards, and Student Testing.

Barber, B. L., & Evans, M. (1992). Policies for reporting test results to parents. *Educational Measurement: Issues and Practice, 11*(1), 15-20.

Croft, C., Gilbert, A., Boyd, S., Burgon, J., Dunn, K., Burgess, L., & Reid, N. (1996). *Assessment Resource Banks in Mathematics and Science. Transition-point assessment – Part 2: Implementation trial.* (Unpublished Report to the Ministry of Education.) Wellington: New Zealand Council for Educational Research.

Ell, F. (2001). *Maths in the New Zealand Curriculum – A concept map of the curriculum document.* (Technical Report 11.) Auckland, NZ: University of Auckland, Project asTTle.

Frary, R. B., Cross, L. H., & Weber, L. J. (1993). Testing and grading practices and opinions of secondary teachers of academic subjects: Implications for instruction in measurement. *Educational Measurement: Issues and Practice, 12*(3), 23–30.

Friedman, S. J., & Frisbie, D. A. (1995). The influence of report cards on the validity of grades reported to parents. *Educational and Psychological Measurement, 55*(1), 5–26.

Glasswell, K., Parr, J., & Aikman, M. (2001). *Development of the asTTle Writing Assessment Scoring Rubrics: Report and Recommendations.* (Technical Report 6.) Auckland, NZ: University of Auckland, Project asTTle.

Guskey, T. R. (1996). Reporting on student learning: Lessons from the past-Prescriptions for the future. In T. R. Guskey (Ed.), *Communicating Student Learning* (pp. 13-24). Alexandria, VA: Association for Supervision and Curriculum Development.

Hambleton, R. K., & Slater, S. C. (1997). *Are NAEP executive summary reports understandable to policy makers and educators?* (CSE Technical Report 430.) Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing, Graduate School of Education & Information Studies, University of California, Los Angeles.

Hattie, J. C., Tognolini, J., Keegan, P. J., & Brown, G. T. L. (in progress). *Standard setting triangulation: Three methods of mapping reading items to curriculum levels* (Technical Report 20.) Auckland, NZ: University of Auckland, Project asTTle.

Henry, G. T. (1993). Using graphical displays for evaluation data. *Evaluation Review, 17*(1), 60–78.

Impara, J. C., Divine, K. P., Bruce, F. A., Liverman, M. R., & Gay, A. (1991). Does interpretive test score information help teachers? *Educational Measurement: Issues and Practice, 10*(4), 16–18.

Jaeger, R. M., Gorney, B., Johnson, R. L., Putnam, S. E., & Williamson, G. (n.d.). *Designing and developing effective school report cards: A research synthesis*. Kalamazoo, Michigan: Center for Research on Educational Accountability and Teacher Evaluation, Western Michigan University.

Jaeger, R. M., & Putnam, S. E. (1994). *Communicating to parents and school board members through school report cards: Effective strategies*. Greensboro, NC: Paper presented to the Annual Meeting of the North Carolina Association for Research in Education.

Lake, K., & Kafka, K. (1996). Reporting methods in grades K-8. In T. R. Guskey (Ed.), *Communicating Student Learning* (pp. 90-118). Alexandria, VA: Association for Supervision and Curriculum Development.

Limbrick, L., Keenan, J., & Girven, A. (2000). *Mapping the English Curriculum.* (Technical Report 4.) Auckland, NZ: University of Auckland, Project asTTle.

Linn, R. L., & Dunbar, S. B. (1992). Issues in the design and reporting of the National Assessment of Educational Progress. *Journal of Educational Measurement, 29*(2), 177–194.

Meagher-Lundberg, P. (2000). *Report on comparison groups/variable for use in analysing assessment results.* (Technical Report 1.) Auckland, NZ: University of Auckland, Project asTTle.

Meagher-Lundberg, P. (2001a). *Report on output reporting design: Focus group 1.* (Technical Report 9.) 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.

Meagher-Lundberg, P., & Brown, G. T. L. (2001a). *Item Signature Study: Report on the Characteristics of Reading Texts and Items from Calibration 1.* (Technical Report 12.) Auckland, NZ: University of Auckland, Project asTTle.

Meagher-Lundberg, P., & Brown, G. T. L. (2001b). *Item Signature Study 2: Report on the Characteristics of Reading Texts and Items from Calibration 2.* (Technical Report 16.) Auckland, NZ: University of Auckland, Project asTTle.

Plake, B. S., Impara, J. C., & Fager, J. J. (1993). Assessment competencies of teachers: A national survey. *Educational Measurement: Issues and Practice, 12*(4), 10–12, 39.

Salvagno, M., & Teglasi, H. (1987). Teacher perceptions of different types of information in psychological reports. *The Journal of School Psychology, 25*, 415–424.

Scriven, M. (1991). Beyond formative and summative evaluation. In M. W. McLaughlin & D. C. Phillips (Eds.), *Evaluation and education: At quarter century.* (Vol. x Part II, pp. 19–64). Chicago, Il.: NSSE.

Spolsky, J. (2001). *User Interface Design for Programmers*. Berkeley, CA: APress LP.

Stiggins, R. J., Frisbie, D. A., & Grisworld, P. A. (1989). Inside high school grading practices: Building a research agenda. *Educational Measurement: Issues and Practice, 8*(2), 5–14.

Wainer, H. (1997). *Visual Revelations: Graphical tales of fate and deception from Napoleon Bonaparte to Ross Perot*. New York, NY: Copernicus Books.

Wainer, H., Hambleton, R. K., & Meara, K. (1999). Alternative displays for communicating NAEP results: A redesign and validity study. *Journal of Educational Measurement, 36*(4), 301–335.

Waltman, K. K., & Frisbie, D. A. (1994). Parents' understanding of their children's report card grades. *Applied Measurement in Education, 7*(3), 223–240.