Designing Alternate Assessment Score Reports that Maximize Instructional Impact

Amy Clark, Meagan Karvonen, Neal Kingston, Gretchen Anderson, and Sheila Wells-Moreaux

University of Kansas

Author Note

Paper presented at the 2015 annual meeting of the National Council on Measurement in Education, Chicago, Illinois. This paper was developed under grant 84.373X100001 from the U.S. Department of Education, Office of Special Education Programs. The views expressed herein are solely those of the author(s), and no official endorsement by the U.S. Department should be inferred. Correspondence concerning this paper should be addressed to Amy Clark, Center for Educational Testing and Evaluation, University of Kansas, akclark@ku.edu.

Abstract

This paper describes an iterative process for designing alternate assessment score reports to be actionable for teachers and parents. The score reports are designed for a new type of alternate assessment for students with significant cognitive disabilities. We present information about multiple rounds of focus groups that informed the development of initial and revised prototypes. We conclude with initial evidence for the potential usefulness of the prototype score report contents when communicating with parents and next steps in the research and development process.

Designing Alternate Assessment Score Reports that Maximize Instructional Impact

Under No Child Left Behind (NCLB), state assessment programs are required to provide "individual student interpretive, descriptive, and diagnostic reports that allow parents, teachers, and principals to understand and address the specific academic needs of students" (NCLB, 2001, §1111[b][3][C][xii]). This reporting requirement encompasses all students, including students with significant cognitive disabilities who take alternate assessments based on alternate achievement standards (AA-AAS). Score reports play an important role in supporting the validity of inferences about scores and the intended uses of those scores. While the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014) require test producers to take steps to promote accurate interpretation and appropriate uses of score reports (Standard 13.5), these steps must be based on an understanding of the assessment's design and intended uses.

The purpose of this paper is to describe an iterative process to designing individual student score reports for a new alternate assessment system called Dynamic Learning Maps (DLM). The paper is organized to follow this iterative cycle. We begin with an historical perspective on AA-AAS scores and score reports, followed by lessons learned from a series of focus groups with parents of students with significant cognitive disabilities. We then present a brief review of principles for the design of effective score reports as applied to the design of the DLM system. Next we describe the initial prototype score reports and data sources that guided refinements to the prototypes. We conclude with initial evidence for the potential usefulness of the prototype score report contents when communicating with parents and next steps in the research and development process.

Historical Perspective of AA-AAS Scoring and Reporting

Alternate assessments for students with significant cognitive disabilities have only existed in most states since they were first required in 2000-01 under IDEA 1997. Early alternate assessments were often based on a combination of functional and academic content, and not all states provided score reports based on these assessments (Thompson & Thurlow, 2001). These assessments became known as alternate assessments based on alternate achievement standards (AA-AAS) after NCLB required the assessments be based on grade level content standards with alternate expectations for achievement. AA-AAS, which have only assessed academic achievement of students with significant cognitive disabilities for about a decade, have some unique challenges with reporting and usability of results, based in part on intended purposes of the assessments, assessment design, and scoring.

By 2006-07, the most frequent purpose for AA-AAS reported by states was to measure student progress or performance on state standards (86%); only 51% indicated AA-AAS assessed students'

individual strengths and weaknesses and 59% reported that a purpose of AA-AAS was to guide classroom instruction (Cameto et al., 2009, Fig. A-2). To be useful when planning instruction, a score report must meet teachers' information needs as they consider how to set and assess progress toward goals; address individual needs; evaluate effectiveness of practice; and assess whether student needs are being met (Marsh, Pane, & Hamilton, 2006).

A prerequisite for effective data-based instructional decision-making is systematic, repeated assessment over the year. Historically, AA-AAS have been portfolios, performance assessments with a limited number of tasks, or rater checklists. Of these, portfolio-based AA-AAS often involve multiple data collection windows but each window is treated as a source of evidence that factors into the overall performance level at the end of the year. The data collected during a window do not necessarily lead directly back into timely instructional decisions, but are saved in a notebook to be submitted at the end of the year.

Another challenge with individual student score reports for AA-AAS is the basis of scores themselves. In 2009, scoring methods for AA-AAS tended to be based on rubrics (67% of states), percent correct (38%), and/or rating scales (19%; Altman et al., 2010). In the same study, rubric criteria included indicators of student performance such as the student's skill, degree of progress, level of assistance required, but some states' scoring systems also included programmatic indicators such as alignment of evidence to academic content standards, participation in general education settings, social relationships, and self-determination. When a student's overall results are based in part on factors other than his or her performance, score reports may be misinterpreted because of the disconnect between what goes into the score and what the reader assumes a score report indicates.

Final performance levels for AA-AAS are often determined by cut scores applied to rubrics or raw scores, as small student populations and limited items have not historically allowed states to apply IRT-based scaled scores. In many states, large percentages of students who take AA-AAS receive scores that are considered proficient or advanced and growth across years is difficult to detect because of the lack of underlying scale, small population sizes, and ceiling effects (Karvonen, Flowers, & Wakeman, 2013). When the status indicator has little meaning (i.e., students with very little evidence of academic knowledge and skills can be "proficient"), score reports that consist primarily of overall status indicators have limited potential for use beyond program and school evaluation.

AA-AAS results that convey little meaning then are not as useful to teachers for educational planning. There is some evidence of their limited utility from research with teachers. For example, when teachers in four states were asked how they know whether students with significant cognitive

disabilities have mastered a skill and how they decide what to teach next, there was little evidence that teachers used assessment results systematically or considered state academic content standards (Karvonen, Wakeman, Moody, & Flowers, 2013). In a sample of IEPs for students who take AA-AAS, fewer than 50% had statements on present levels of academic and functional performance that included criterion-based evidence of the student's prior performance in reading or math (Karvonen, Rao, & Morgan, 2010). Because the shift to grade level-aligned academic content was a dramatic change in the expectations for students with significant disabilities, and because teachers play such a central role in the development (portfolio) or delivery (performance assessment) of AA-AAS, teacher beliefs about the relevance of academic curriculum and value of AA-AAS influence their use of scores. When teachers believe AA-AAS impacts instruction, their students have a higher likelihood of AA-AAS proficiency (Karvonen et al., 2013). Without compelling and relevant individual student score reports, based on assessments designed and scored to provide meaningful data, AA-AAS results have little chance of informing decisions that improve students' educational opportunity.

Parent Perspectives

In 2013 DLM contracted with The Arc (a national advocacy organization for individuals with intellectual and developmental disabilities) to conduct focus groups with parents of children with disabilities. Five focus groups were conducted across five states, with a total of 44 participants. Follow-up interviews were conducted by phone with seven individuals. While the study included several research questions about topics including academic expectations, post-secondary goals, and transition plans (Nitsch, 2013), in this paper we summarize the key findings around (1) parent understanding of alternate assessments and (2) parent need for information about student performance. Participants were asked about their current alternate assessment process and reports as well as their ideas for how information could be presented to maximize its use in helping their children attain goals.

In these focus groups, parents rated themselves as having relatively little knowledge of AA-AAS and some indicated they had not received score reports from their schools. Parents tended to perceive the purpose of AA-AAS as to fulfill a legislative mandate and to drive decisions about the school (including teacher evaluation and determination of resources) rather than to provide information about their child or measure things relevant to their learning (Nitsch, 2013). Concerns about the information parents received on AA-AAS results included lack of understanding of how scores were determined or how the content was related to academic content standards, unfamiliar terminology, a focus on deficits more so than progress, and lack of information about how results could be used to change instruction or provide different supports to their child.

When asked what information they needed about their student's results from AA-AAS, parents suggested ideas such as:

- Providing a context for understanding scores
- Explaining the purpose of the test
- Connecting results to planning new educational goals
- Identifying areas of growth as well as areas that need more progress (Nitsch, 2013, pp. 29, 31)

Parents also suggested resources should help them understand what the assessment looks like, how the contents were decided upon, how the results fit into accountability systems, and what actions schools would take based on the results.

Report Design Principles and the DLM Assessment System

General Principles for Report Design

An actionable report facilitates decision-making, which means that the report contents must be interpretable and as well as useful. At a minimum, for accurate interpretation, a score report must include (a) what the test covers, (b) what scores mean, (c) score precision, and (d) how the scores will be used (AERA, et al., 2014, Standard 6.10). In addition to the professional standards, research-based practices identify report attributes believed to facilitate report interpretation and use (Blackwell, 2012; Goodman & Hambleton, 2004). Effective student reports are described as "clear, concise, and visually attractive; also should include easy-to-read text that supports and improves the interpretation of charts and tables; … statistical jargon should be avoided; data should be grouped in meaningful ways" (Goodman and Hambleton, 2004, p. 64). More recently, Zenisky and Hambleton (2012) have expanded the best practice for effective score reports to include appropriate information about subdomains of the test, skills and abilities required for the different identified levels of proficiency, and useful comparisons with the proper population.

Beyond general best practices, it is important to remember that score report contents must target the intended audience in order to support intended uses (Ryan, 2006). However, matching a report to the intended audience is more than simply providing the right data. Other attributes, such as audience reading skills and behaviors, will also affect report utility. For example, approximately 40% of parents will struggle with using a score report due to basic literacy deficits, regardless of any issues related to communicating psychometric data (Kutner, et al., 2007). When using a report, many parents simply review the overall performance category and, based on this information, decide what other parts of the report they will read, if any (NEGP, 1998; Trout and Hyde, 2006). Unfortunately, few reports provide a clear link between a score and any information needed to interpret the score (Anderson,

2014). It is likely that information included in sidebars, cover letters, and separate guides contribute very little to score interpretation because they are outside the score report itself.

DLM System Design

While the literature provides guidance for the development of score reports for large-scale assessment programs, some elements of DLM score reports will necessarily be different because of the design of the system and the way scores are generated. The Dynamic Learning Maps (DLM) system is designed to assess a student's learning throughout the year. Assessments take the form of testlets, each of which contains an engagement activity and 3-8 items. Testlets are available for teachers to use on an instructionally embedded basis through most of the year. A spring testing window allows for full coverage of the blueprint or a resampling of content covered throughout the year, depending on which testing model the state uses.

DLM assessments are based on large, fine-grained learning maps made up of multiple pathways. Nodes in the map represent knowledge, skills, and understandings in English language arts and mathematics as well as foundational areas that are needed for access to academics. Nodes are linked to Essential Elements (EE), the grade-level expectations for students with significant cognitive disabilities. Testlets are available at five linkage levels per EE. For each EE on the blueprint, one or more nodes aligned with the expectation in the EE are grouped together to form the target linkage level. Groups of nodes are also identified in three areas that come before the target (initial precursor, distal precursor, and proximal precursor) and one that stretches past the target (called the successor). Students typically take one testlet per EE, at one linkage level, the assignment of which is based on prior information about the student and modeling of the learning map.

Essential Elements are organized further into claims and conceptual areas. The DLM claims are overt statements about what is intended for students to learn and what the DLM assessment will measure. The claims encompass the portion of the learning maps in ELA and mathematics that are connected to the DLM Essential Elements. Subareas of the Claims, called Conceptual Areas, connect the learning map to the overall Claims and identify large areas of conceptually related skills in the maps. Conceptual Areas are areas of the learning maps within Claims organized around common cognitive processes. For example, one of the English language arts claims is that "students will comprehend text in increasingly complex ways." One of the conceptual areas for that claim is "integrating ideas and information from text."

DLM uses cognitive diagnostic modeling to calculate the student's probability of mastery for every node in the learning map. Final assessment results are based on mastery classification (whether a

student demonstrated mastery of the node or not) rather than a score on a latent trait. Because results are based on dichotomous classifications at the node level, there is no continuous scale score. DLM scores are built up from the node, rather than starting with a scale score that is then broken down into subdomains. All DLM score reporting is based on the node level mastery classifications, summarized across various levels of aggregation to make statements about each student's knowledge, skills, and ability. The highest level of aggregation is the performance level. Performance levels are defined through a standard setting process based on profiles of linkage levels mastered. The majority of states in the DLM consortium have adopted 4 performance levels with the following labels: emerging, approaching the target, at target, and advanced.

As further grounding for the next section on score report prototypes, we conclude with some information about intended uses of results and ways in which the score reports are expected to provide validity evidence. The DLM consortium has identified several intended uses of results from DLM assessments:

1. Reporting achievement and growth within the taught content aligned to grade-level content standards, to a variety of audiences, including educators and parents
2. Inclusion in state accountability models to evaluate school and district performance
3. Planning instructional priorities and program improvement for following school year

Results of instructionally embedded assessments are also appropriately used for instructional planning, monitoring, and adjustment.

Specific to score reporting, the validity argument for DLM includes the following propositions about score reports and score interpretation:

- Score reports are useful and provide relevant information for teachers
- Scores are only interpreted and used for those purposes supported by the validity argument
- Teachers can use score reports to inform instructional choices and goal setting

Data collected during the score report design stage is one source of evidence to be used when evaluating the validity argument.

Prototype Score Reports

Going beyond broad performance level descriptors, DLM reports are designed to provide actionable information to guide instructional decisions while also being appropriate for use for accountability purposes. Reports must support teachers and parents in interpreting results that differ from any they have likely seen before (linkage level mastery) and guard against potential misinterpretations or misuses of the results.

Three score reports have been developed at the individual student level: 1) a progress report, 2) an end of year learning profile, and 3) an end of year performance profile. The prototypes included in the appendices to this paper are the final versions for the 2015-16 school year. Each prototype was first developed by staff based on research literature and the initial round of focus groups described earlier. Final prototypes reflect feedback obtained during the second set of focus groups.

The **progress report** (see prototype excerpt in Appendix A) is available to teachers on demand during the school year and reflects all testing and instructional plans to date. The report is organized first by conceptual area then by Essential Element. For each Essential Element, the five possible linkage levels are included as cells that state the linkage level descriptor. Shading is used to distinguish between levels that the student mastered, levels that are current instructional goals, and levels that have not been taught or assessed to date. For levels that have been mastered by the student, the date of mastery is included in the cell. By including all linkage levels, regardless of current instructional goal or mastery, the progress report demonstrates the student's progress towards the grade level expectation. The progress report is intended for the teacher to discuss with the parent, rather than as a stand-alone report that is sent home directly to the parent.

The end of year **learning profile** (see prototype excerpt in Appendix B) is very similar in structure and content to the progress report. The report is also organized by conceptual area and Essential Element, but rather than being available on demand during the year, the learning profile provides a summary of performance at the end of the year. Shading is used to distinguish between levels the student mastered, levels assessed but not mastered, and levels not assessed that year. Date of mastery is not included on the end of year learning profile. This report is intended to be sent directly to the parent at the end of the academic year.

The **end of year performance profile** (see prototype in Appendix C) provides a higher-level summary of student performance than the learning profile. Rather than reporting information by Essential Element nested within conceptual area, this report provides only a high level summary of student mastery at the conceptual area. The performance profile contains two main sections. The first section (Overall Results) includes a text summary of the student's performance, including the total number of Essential Elements mastered during the year and the student's final performance level. The second section (Conceptual Areas) provides a detailed summary of student performance for each conceptual area. Graphics demonstrate the percent of skills mastered by conceptual area. This value is calculated as the number of linkage levels mastered (as reported on the learning profile) out of the total

number of linkage levels possible for the grade and content area. This section also includes general statements of performance typical of students at the same performance level as the student.

All three student-level reports depict "student mastery." Mastery here is defined on a three-tiered system. The first level of mastery is at the node level. The use of cognitive diagnostic modeling allows for the calculation of the student's probability of mastery for every node in the learning map. As part of the standard setting process, a threshold for node mastery is set (e.g. 0.8). If the student's probability of node mastery is greater than this threshold, the student is considered a master of the node. If the student's probability of mastery is below the threshold, the student is considered a non-master of the node. Node level mastery is not included on the individual student reports.

The next tier of aggregation occurs at the linkage level. In many instances, more than one node is assessed at a linkage level. To set a threshold of mastery for the linkage level, the proportion of nodes mastered at the linkage level must also be set as part of the standard setting process (e.g. 0.75). If the student's proportion of nodes mastered is greater than this threshold, the student is considered a master of the linkage level. Using the 0.75 as an example threshold value, if there are four nodes assessed at the linkage level, the student would need to have mastered at least three out of four nodes to be considered a master of the linkage level. If the student's proportion of nodes mastered is below the threshold, the student is considered a non-master of the linkage level. Linkage level mastery is represented on the progress report and learning profile as the shaded boxes, which indicate the linkage levels mastered by the student.

The third tier of aggregation summarizes linkage level mastery to arrive at the performance level. Again, the standard setting process must be conducted to specify the linkage level mastery that is associated with each performance level. The student's final performance level is displayed on the performance profile along with the descriptors for each performance level.

Figure 1 provides an overview of the levels of aggregation and which score report summarizes the student's performance at each level.

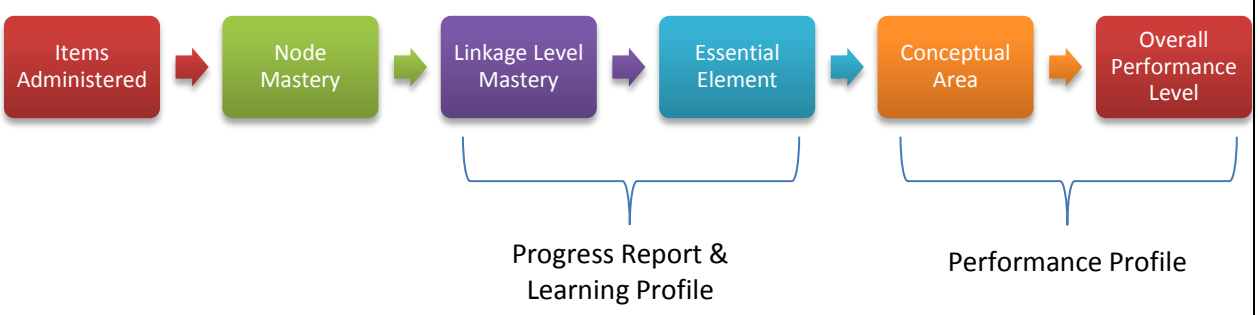


Figure 1. Steps in process of aggregating mastery information and primary location of information in reports

Feedback on Prototypes

To obtain feedback on initial prototypes of DLM score reports, four focus groups were held in the spring of 2014 with parents of students with significant cognitive disabilities. An additional focus group was held with educators. Two full prototype sets were provided for the first two focus groups. There was one report for each subject. The two examples had contrasting patterns of student performance: low mastery but high growth, and high mastery but low growth. Two variants of the learning profile were presented to allow easy comparisons of the interpretability and utility of each. A summary of feedback and changes to each report between initial prototype and the final versions in the appendices is provided below.

Progress Report

The initial prototype of the progress report listed EEs, grouped by conceptual area, and used green shading to indicate mastery and yellow shading to represent current instructional targets. The names of linkage levels were not included in this report, but the grade level target was identified with a graphic (see Figure 2).

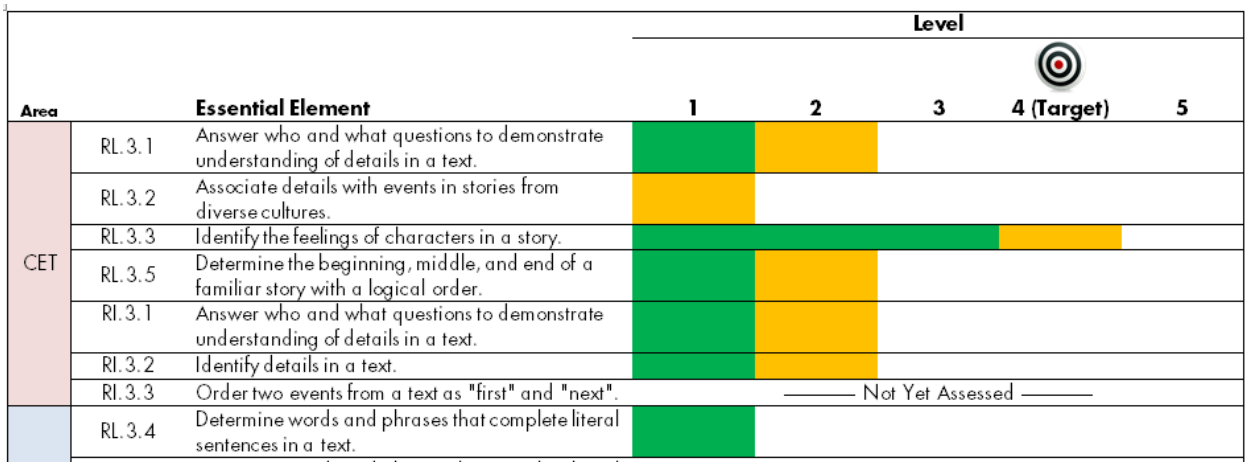


Figure 2. Initial prototype progress report.

Parents indicated that the terminology (e.g., “Essential Element” or “Conceptual Area”) was unfamiliar and difficult to understand. The use of acronyms for conceptual area, the key for which was located in the report footer, was also a source of confusion. The shading of linkage levels, originally intended to provide information approximating a bar graph, was not informative enough. The relationship of current areas of instruction (yellow cells) to the column labeled “Target” also caused confusion:

> Parents did not know whether the targets were for their own child, for all students in the alternate, or for all students at the grade level in and out of the alternate. Some parents read the report as indicating that some students’ individual goals will be set below the target, and did not understand how this determination was made or why it was acceptable for a student to have lower goals set if a target was indicated. Parents said they would like an explanation of what the target means, who it applies to, and how these targets and individual student goals are determined. (Nitsch, 2014, p. 3)

Some of this confusion was addressed by expanding the introductory text at the beginning of the report and adding text to describe the target linkage level for each EE so the relationship between the EE, target, and current level was clearer. The versions of introductory text in the initial and final prototypes are reproduced in Table 1.

Table 1

Initial and Final Versions of Introductory Text

Version	Introductory Text
Initial	Susie's current performance in 3rd grade English language arts Essential Elements is summarized below. This information is based on all of the DLM tests she has taken between the beginning of the school year and **January 23, 2015**. The target level is the grade level expectation for students to have proficient understanding and application of the Essential Element.
Final	Susie's current performance in 3rd grade English language arts Essential Elements is summarized below. This information is based on all of the DLM tests she has taken between the beginning of the school year and **January 23, 2015**. In order to master an Essential Element, a student must master a series of skills leading up to the specific skill identified in the Essential Element. This table describes what skills your child demonstrated in the assessment, and how those skills compare to grade level expectations. This report does not show progress on all of Susie's instructional goals. She may be taught other academic concepts that have not yet been tested. This report does not show progress on her IEP goals.

Information about student progress, areas of strength and weakness were highly valued. Parents wanted to know where child was stalled in his or her learning, or where s/he was on the path toward a larger goal. At participants' recommendation, we added mastery dates to the mastered linkage levels (see Appendix A) so they could think about the rate at which their child might be expected to reach the next level.

Learning Profile

The learning profile prototype is very similar to the progress report but does not include language about instructional goals. Since the learning profile is part of a summative report delivered at the end of the year, the first prototype showed shading for all mastered levels (see Figure 3). One version also showed linkage level descriptors for the highest level mastered.

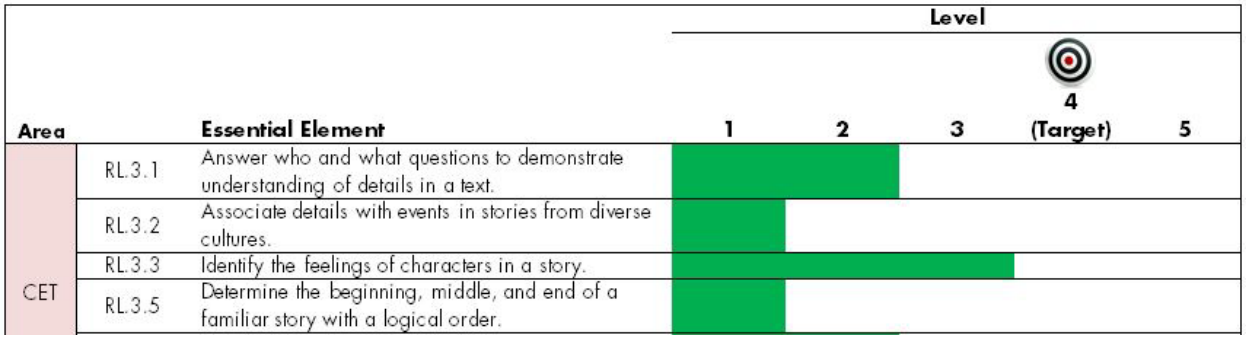


Figure 3. Initial prototype learning profile #1.

Many of the lessons learned about the progress report transferred to the learning profile. Parents and educators valued the inclusion of linkage level descriptors. In an attempt to balance amount of information with meeting requests for more information, three focus groups examined a version of the learning profile with no linkage level descriptors and a version with descriptors at the target and the highest level mastered.

			Level				
Area		**Essential Element**	**1**	**2**	**3**	**4 (Target)**	**5**
CE	RL.3.1	Answer who and what questions to demonstrate understanding of details in a text.		Identify familiar people, objects, places, or events		Answer who and what questions about story details	
	RL.3.2	Associate details with events in stories from diverse cultures.	Seek absent objects			Associate details with events in diverse stories	
CU	RL.3.4	Determine words and phrases that complete literal sentences in a text.	Attend to object characteristics			Complete the meaning of a sentence	

Figure 4. Initial prototype learning profile #2.

Parents preferred the second report and indicated it would be useful for planning meetings and conversations with teachers. They also described it as visually appealing and easy to understand. Participants noted that the example in Figure 2 helped the reader see discrepancies between actual mastery and the grade level expectation (target), but that the report had limited value for instructional planning when mastery was low; the reader would not know what steps came between the highest level mastered and the target. Despite staff concerns about the volume of information on the page, the final prototype included linkage level descriptors at all levels for all EEs on the report (Appendix B).

End of Year Performance Profile

To support the original goal of developing reports that fit the design of the DLM system, we initially developed performance profiles to include information about fine-grained, with and across year growth in linkage level mastery. The initial prototype contained information about overall performance and a parallel description of growth (Figure 5). Each included norm-referenced interpretations with comparisons to the district, state, and nationwide DLM student populations. Results were expressed in terms of levels mastered (status) and EEs in which the student showed expected growth.

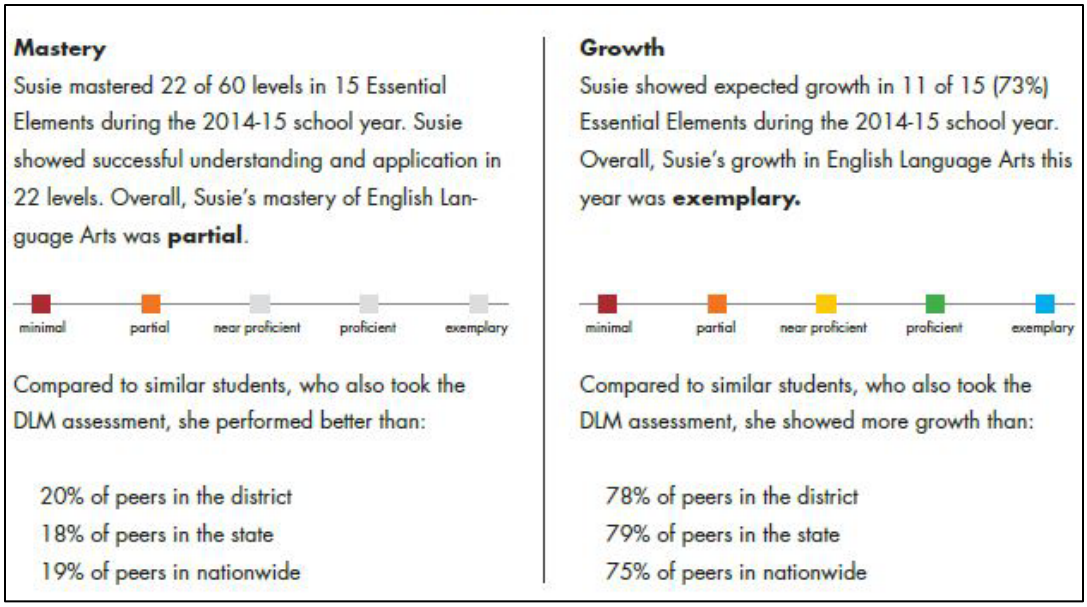


Figure 5. Initial prototype of performance profile – overall results.

Similar to many existing score reports for large-scale assessment, we also included information about linkage level mastery by conceptual area, to give the reader a sense of the student's performance at a grain size between the overall performance in the subject and EE-specific information (Figure 6).

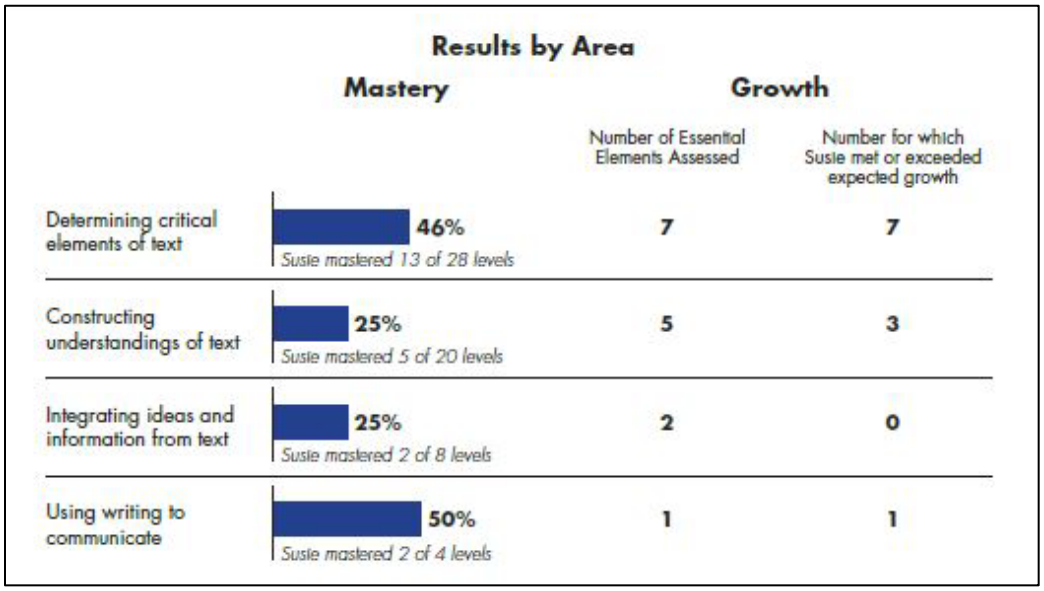


Figure 6. Initial prototype of performance profile – results by conceptual area.

Parents responded positively to these reports and indicated they were more helpful than their states' current AA-AAS reports. Parents also responded positively to the norm-referenced information, although they needed clarification on whether the comparison was to all students who took DLM or just those with similar abilities. They understood the contents of these reports more quickly than the

learning profiles, but believed the learning profiles better communicated about their child's achievements. They understood the bar graphs representing mastery in a conceptual area better than they understood the shaded boxes indicating overall performance level. The presentation of graphs for mastery next to numbers for growth introduced some confusion, as parents had to work to connect this information to other parts of the report. As the consortium states decided to move away from reporting growth on the individual student score report, that information was omitted. Educators also expressed strong concerns about the inclusion of normative information in the reports, especially for the parents whose students would not perform well in comparison to their peers. This interpretation is not on the 2014-15 prototype because we do not yet have the data to support norm-referenced interpretations. This part of the report will be revisited once data become available and stakeholders provide additional input.

A third part of the performance profile is the narrative description of the student's knowledge and skills (see Appendix C). Realizing that some parents will not read the amount of detail in the learning profile, this section of the performance profile is designed to provide information about mastery at the conceptual area level. This section of the report was relatively unchanged from earlier prototypes to later ones. Across focus groups, this section was viewed as helpful. Parents appreciated the focus on the student's strengths (what was mastered) rather than deficits (what was not mastered). Educators noted that the statements would integrate well into the statements on present levels of academic performance required on IEPs.

Early Evidence on Interpretability

In spring 2015, five current educators were introduced to DLM concepts needed to understand score reports (e.g., EE, mastery) and provided with a sample Learning Profile and Performance Profile. After being given a few minutes to review the reports without further explanation, they had an opportunity to ask questions. They then were asked to imagine themselves using the report to talk to the parent of the student whose sample report they reviewed. They wrote their comments to parents before discussing the report as a group.

Despite having access to overall performance and performance by conceptual area as well as more fine-grained information about linkage level mastery, all five focused their interpretive statements on the linkage level descriptions. None mentioned overall performance levels or quantities of linkage levels mastered. All five led with statements about what the student could do. Three of the five also mentioned areas for future instructional focus, based on descriptions of linkage levels not yet mastered. For example: "Susie is good in identifying familiar objects and people but is still learning to understand

words…We need to work on identifying words and noticing new things." Two participants used vocabulary specific to DLM, such as "node" or "linkage level." One participant expressed statements of relative strength and weakness in terms of conceptual areas: "[the student is] best at determining critical elements of a text. The next place we want to work on improving would be constructing understandings of text."

Conclusions and Next Steps

Assessments are more likely to fail in their intended purposes if the resulting information is uninterpretable by the intended report user. Design research is a necessary step for achieving interpretable and useful score reports. This paper describes several steps in the design process and lessons learned on the path toward final prototype score reports for the consortium's first year of operational testing.

We expect these reports will evolve in future years as we learn more about widespread teacher and parent use of the reports. The parents and educators who participated in the focus groups described in this paper are by no means representative of the population that will interact with DLM score reports. The parents were extremely well informed, and many of them also served as parent advocates or had education backgrounds. While the parent advocates considered the needs of parents they worked with when providing feedback on the prototypes, it is likely that the first year's reports will have language that is still challenging and technical for many parents, especially since they are also transitioning to DLM assessments for the first time and have no prior experience with this assessment system. Future research should also help us evaluate the balance between quantity of information and interpretability. The current prototypes provide a wealth of information for parents who want to dig deep into an understanding of their child's learning. Hopefully we have avoided the unintended consequence of causing other parents to find the report so overwhelming as to be a barrier to their understanding of their child's achievement.

These report prototypes are finalized for 2015-16 and are now ready for the next two phases of the design process, usability testing and heuristics. Usability testing focuses on measuring a report's capacity to meet its intended purpose. Paraphrase testing is one method we intend to use to evaluate document readability and interpretability. In order to evaluate the report across a wide range of likely report users, paraphrase testing will be conducted with new and experienced teachers and with parents, including parents who speak English as a second language. To address the expectation that parents and teachers use the reports when making educational decisions for a student, a second usability test will

include paired, or co-, discovery. Paired discovery involves observing two users working together to successfully achieve a task using information provided within the report, while thinking aloud.

In addition to ongoing research and development, a more immediate step we will take to support operational assessment is to develop resources to support parent and educator interpretation and use of score reports. Eventually, DLM reports will be delivered electronically so the reader will have on-demand access to information that helps them understand what they are reading. In the short term, reports are static and support for interpretation will come in other forms. As noted in the first round of parent focus groups (Nitsch, 2013), parents want information beyond what fits in a score report. Because this information might not be visible and easily connected to score reports, extra care must be taken to connect explicitly report content to the supplemental resources. Parents will not intuitively know what contextual information is necessary in order to interpret a report accurately and are likely to assume greater understanding than is warranted.

While focus group participants responded positively to the prototypes, those parents still expressed skepticism about teachers' abilities to communicate about the report and their time to do so. The supporting resources, including interpretation guides and a professional development module, will be designed to help both groups be best prepared with knowledge about the DLM system in general and an understanding of the score report contents more specifically so that their limited time may be spent discussing action steps rather than educating parents about the meaning of report contents. More time spent on meaningful conversations about student learning will further the goal of high expectations, and goal attainment, for students with significant cognitive disabilities.

References

Altman, J. R., Lazarus, S. S., Quenemoen, R. F., Kearns, J., Quenemoen, M., & Thurlow, M. L. (2010). *2009 survey of states: Accomplishments and new issues at the end of a decade of change*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved from http://www.cehd.umn.edu/NCEO/OnlinePubs/StateReports/2009_survey_of_states.htm

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, & Joint Committee on Standards for Educational Psychological Testing. (2014). *Standards for Educational and Psychological Testing*: AERA.

Anderson, G. (2014). A language parents understand: Reporting and adult functional literacy research. Presented at the annual meeting of the National Council on Measurement in Education, Philadelphia, PA.

Blackwell, W. H., III. (2012). *Making alternate assessment score reports a meaningful tool for parents.* (Ed.D. 3520230), Boston University, United States -- Massachusetts. Retrieved from http://search.proquest.com/docview/1033079548?accountid=14556 ProQuest Dissertations & Theses (PQDT) database.

Cameto, R., Knokey, A.-M., Nagle, K., Sanford, C., Blackorby, J., Sinclair, B., and Riley, D. (2009). *National Profile on Alternate Assessments Based on Alternate Achievement Standards. A Report From the National Study on Alternate Assessments* (NCSER 2009-3014). Menlo Park, CA: SRI International. Retrieved from http://ies.ed.gov/ncser/pdf/20093014.pdf

Goodman, D. P., & Hambleton, R. K. (2004). Student Test Score Reports and Interpretive Guides: Review of Current Practices and Suggestions for Future Research. *Applied Measurement in Education, 17*(2), 145-220.

Karvonen, M., Flowers, C., & Wakeman, S. (2013, April). *An exploration of methods for measuring academic growth for students with significant cognitive disabilities*. Paper presented at the 2013 annual meeting of the American Educational Research Association.

Karvonen, M., Rao, V., & Morgan, G. (2010, May). *Relationships between curricular priorities, instructional practices, and state-wide assessment scores for students eligible for AA-MAS.* Paper presented at the annual meeting of the American Educational Research Association, Denver, CO.

Karvonen, M., Wakeman, S., Moody, S., & Flowers, C. (2013). The relationship of teachers' instructional decisions and beliefs about alternate assessments to student achievement. *Exceptionality*, 21, 238-252. doi: 10.1080/09362835.2012.747184

Kutner, M., Greenberg, E., Jin, Y., Boyle, B., Hsu, Y. C., & Dunleavy, E. (2007). Literacy in Everyday Life: Results from the 2003 National Assessment of Adult Literacy. NCES 2007-490. *National Center for Education Statistics*.

Marsh, J. A., Pane, J. F., & Hamilton, L. S. (2006). Making sense of data-driven decision making in education (Paper No. OP-170-EDU). Retrieved from www.rand.org

National Education Goals Panel (NEGP). (1998). *Talking about tests; An idea book for state leaders.* Washington, DC: US Government Printing Office.

Nitsch, C. (2013). *Dynamic Learning Maps: The Arc parent focus groups*. Unpublished manuscript. Washington, DC: The Arc.

Nitsch, C. (2014, September 14). *Parent focus groups: Feedback on mock score reports.* Unpublished manuscript. Washington, DC: The Arc.

Quenemoen, R., Rigney, S., & Thurlow, M. (2002). *Use of alternate assessment results in reporting and accountability systems: Conditions for use based on research and practice* (Synthesis Report 43). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved from http://education.umn.edu/NCEO/OnlinePubs/Synthesis43.html

Thompson, S., & Thurlow, M. (2001). *2001 State special education outcomes: A report on state activities at the beginning of a new decade*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved from http://education.umn.edu/NCEO/OnlinePubs/2001StateReport.html

Trout, D. L, & Hyde, E. (2006). *Developing score reports for statewide assessments that are valued and used: Feedback from K-12 stakeholders.* Presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Zenisky, A. L., & Hambleton, R. K. (2012). Developing test score reports that work: The process and best practices for effective communication. *Educational Measurement: Issues and Practice*, *31*(2), 21-26.

Susie's current performance in 3rd grade English language arts Essential Elements is summarized below. This information is based on all of the DLM tests she has taken between the beginning of the school year and **January 23, 2015**.

In order to master an Essential Element, a student must master a series of skills leading up to the specific skill identified in the Essential Element. This table describes what skills your child demonstrated in the assessment, and how those skills compare to grade level expectations.

This report does not show progress on all of Susie's instructional goals. She may be taught other academic concepts that have not yet been tested. This report does not show progress on her IEP goals.

			Level Mastery				
Area	**Essential Element**	**Grade Level Expectation**	**1**	**2**	**3**	**4 (Target)**	**5**
Determining Critical Elements of Text	RL.3.1	Answer who and what questions to demonstrate understanding of details in a text.	Attend to object characteristics **10/2/14**	Identify familiar people, objects, places, and events	Answer who and what questions and identify details in a familiar story	Answer who and what questions about story details	Answer who, what, when, and where questions about story details
	RL.3.2	Associate details with events in stories from diverse cultures.	Seek absent objects **10/12/14**	Identify familiar people, objects, places, or events	Associate details with events in a familiar story	Associate details with events in diverse stories	Recount diverse stories with key details
	RL.3.3	Identify the feelings of characters in a story.	Identify feeling states in self	Identify feeling words **11/7/14**	Identify the feelings of characters in familiar stories	Identify the feelings of characters in a story	Identify character feelings and relate to actions
	RL.3.5	Determine the beginning, middle, and end of a familiar story with a logical order.	Express interest in book sharing **8/12/14**	Differentiate between text and pictures	Identify details and beginning and end of a story	Determine the beginning, middle, and end of a familiar story with a logical order	Identify beginning and end of a story

Individual Student Year-End Report
Learning Profile

DYNAMIC™ LEARNING MAPS

NAME: Susie Smith | SCHOOL: DLM School | YEAR: 2014 – 15
SUBJECT: English Language Arts | DISTRICT: DLM District | GRADE: 3
REPORT DATE: 06-10-2015 | STATE: DLM State | ID: 08691

Susie's performance in 3rd grade English Language Arts Essential Elements is summarized below. This information is based on all of the DLM tests she took during the 2014-15 school year.

In order to master an Essential Element, a student must master a series of skills leading up to the specific skill identified in the Essential Element. This table describes what skills your child demonstrated in the assessment, and how those skills compare to grade level expectations.

Green shading shows levels she mastered this year. Blue shading shows levels assessed but not mastered this year.

Area	Essential Element	Level Mastery 1	2	3	4 (Target)	5	Did student grow by at least one level this year?
Determining Critical Elements of Text	RL.3.1	Attend to object characteristics	Identify familiar people, objects, places, and events	Answer who and what questions and identify details in a familiar story	Answer who and what questions about story details	Answer who, what, when, and where questions about story details	YES
	RL.3.2	Seek absent objects	Identify familiar people, objects, places, and events	Associate details with events in a familiar story	Associate details with events in diverse stories	Recount diverse stories with key details	YES
	RL.3.3	Identify feeling states in self	Identify feeling words	Identify the feelings of characters in familiar stories	Identify the feelings of characters in a story	Identify character feelings and relate to actions	YES
	RL.3.5	Express interest in book sharing	Differentiate between text and pictures	Identify details and beginning and end of a familiar story	Determine the beginning, middle, and end of a familiar story with a logical order	Identify beginning and end of a story	NO

Levels mastered this year Levels assessed but not mastered this year

Individual Student Year-End Report

Performance Profile

DYNAMIC™ LEARNING MAPS

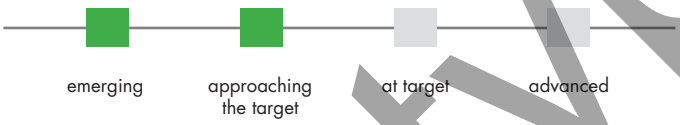
NAME: Susie Smith
SUBJECT: English Language Arts
REPORT DATE: 06-10-2015

SCHOOL: DLM School
DISTRICT: DLM District
STATE: DLM State

YEAR: 2014—15
GRADE: 3
STATE ID: 08691

Overall Results

Grade 3 English language arts allows students to show their achievement in 85 skills related to 17 Essential Elements. Susie has mastered 32 of those 85 skills during the 2014-15 school year. Overall, Susie's mastery of English language arts fell into the second of four performance categories: **approaching the target**. The specific skills Susie has and has not mastered can be found in her Learning Profile.



EMERGING:	The student demonstrates **emerging** understanding of and ability to apply content knowledge and skills represented by the Essential Elements.
APPROACHING TARGET:	The student's understanding of and ability to apply targeted content knowledge and skills represented by the Essential Elements is **approaching the target**.
AT TARGET:	The student's understanding of and ability to apply content knowledge and skills represented by the Essential Elements is **at target**.
ADVANCED:	The student demonstrates **advanced** understanding of and ability to apply targeted content knowledge and skills represented by the Essential Elements.

Conceptual Areas

Determining critical elements of text
43%
Susie mastered 17 of 40 skills

Constructing understandings of text
28%
Susie mastered 7 of 25 skills

Integrating ideas and information from text
40%
Susie mastered 4 of 10 skills

Using writing to communicate
40%
Susie mastered 4 of 10 skills

Individual Student Year-End Report
Performance Profile

NAME: Susie Smith | SCHOOL: DLM School | YEAR: 2014—15
SUBJECT: English Language Arts | DISTRICT: DLM District | GRADE: 3
REPORT DATE: 06-10-2015 | STATE: DLM State | STATE ID: 08691

Performance Profile Continued

More information about Susie's performance on each Essential Element, that make up the Conceptual Areas, is located in her Learning Profile.

Determining Critical Elements of Text

Susie is interested in shared reading. Susie understands actions that are part of routines familiar to her. Susie understands that words have meanings that relate to people and objects around her. Susie can identify characters' feelings and illustrations in familiar texts.

Constructing Understandings of Text

Susie has shown that she can identify objects based on words that describe objects. Susie notices new things in her environment. Susie understands some feeling words.

Integrating Ideas and Information from Text

Susie can identify familiar people, places objects and events.

Using Writing to Communicate

Susie has shown interest in making marks on paper in order to write.