Development and Evaluation of Diagnostic Score Reports for an Alternate Assessment System

Meagan Karvonen, Amy K. Clark, Russell Swinburne Romine, Neal Kingston

University of Kansas

Paper presented at the 2019 annual meeting of the American Educational Research Association, Toronto, Ontario. Some studies described in this paper were conducted 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 of Education should be inferred. Correspondence concerning this paper should be addressed to Meagan Karvonen, ATLAS, University of Kansas, karvonen@ku.edu. Do not redistribute this paper without permission of the authors. The authors wish to thank Liz Kavitsky for her assistance in preparing this manuscript.

Abstract

Actionable score reports facilitate instructional decision-making, which means that contents of the report must be interpretable and useful. The *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association, & National Council on Measurement in Education, 2014) indicate that score reports should include information on (a) what the test covers, (b) what the results mean, (c) precision of measurement, and (d) how results should be used. There are historic challenges to accomplishing these goals for alternate assessments based on alternate academic achievement standards (AA-AAS). This paper describes a process for iterative development and evaluation of diagnostic score reports for a large-scale alternate assessment system, situated in research and guidance regarding best practices for reporting. The paper concludes with lessons learned and potential future directions.

Development and Evaluation of Diagnostic Score Reports for an Alternate Assessment System

The Every Student Succeeds Act (ESSA) requires that statewide academic achievement assessment programs

> produce individual student interpretive, descriptive, and diagnostic reports…regarding achievement on such assessments that allow parents, teachers, principals, and other school leaders to understand and address the specific academic needs of students, …in an understandable and uniform format, and to the extent practicable, in a language that parents can understand (1111)(b)(2)(B)(x)

There is a considerable literature base on standards, procedures, and empirical evidence on the design and use of score reports. The *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association, & National Council on Measurement in Education, 2014) indicate that score reports should include information on (a) what the test covers, (b) what the results mean, (c) precision of measurement, and (d) how results should be used. Literature exists that informs specific design features (e.g., Zenisky & Hambleton, 2012) and decisions to make when tailoring report contents so they communicate clearly to a range of audiences (Zapata-Rivera & Katz, 2014). Deng and Yoo (2009) compiled an extensive annotated bibliography designed as an aid for score report developers, with sources ranging from general reporting guidelines to scores and data displays and including links to sample reports.

The bulk of research on score report design for large-scale academic assessments is based on assessments for which IRT-based unidimensional scale scores are the basis of reporting (Leighton & Gierl, 2010). The literature base on score report design based on fine-grained diagnostic assessment systems, often scored using diagnostic classification models, is scant in comparison. Leighton & Gierl (2010) illustrated a method for reporting results for one domain within mathematics based on an attribute hierarchy method. Sinharay, Puhan, and Haberman (2010) analyzed the literature and raised concerns about the appropriateness of subscore reporting if subscore domains were not unique. Clark and Kingston (2019; paper #1 in this symposium) detailed challenges in communicating results from diagnostic assessments in large-scale systems, such as teacher misinterpretation of the meaning of "mastery", how mastery decisions are aggregated to report overall performance, and the need for quality resources to support teacher interpretation and use.

Aside from specific recommendations for score report design elements, one useful resource is Hambleton and Zenisky's (2013) 7-step process for creating score reports. Their recommendations align steps with phases in the development process, and they offer guidance about issues to consider at each step and stakeholders to involve. The final step in their recommended process, monitoring score report use and inferences being made, brings fewer recommendations for systematic inquiry; the reader is encouraged to "consider ways to connect with intended users" (p. 491). O'Leary, Hattie, and Griffin (2017) extend beyond that guidance and argue for the importance of gathering evidence to evaluate end users' actual interpretation and use of operational score reports.

Alternate Assessment Score Reports

Despite the evidence base on score report design, score reporting practices for large-scale alternate assessments for students with the most significant cognitive disabilities have unique challenges. These assessments have only existed in most states since they were first required in 2000-01 under IDEA 1997 and became known as alternate assessments based on alternate achievement standards (AA-AAS) after NCLB (2002) required the assessments be based on grade level content standards with alternate expectations for achievement. AA-AAS have some unique challenges with reporting and usability of results, based in part on intended purposes of the assessments, assessment design, and scoring. In a survey based on states' 2006-07 AA-AAS, the most frequent purpose states reported for the assessments was to measure student progress or performance on state standards (86%); only 51% indicated their 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).

Historically, AA-AAS scores themselves have not supported instructional uses. 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 received scores that are considered proficient or advanced; growth across years is difficult to detect because of the lack of underlying scale, small population sizes, and ceiling effects (Karvonen, Flowers, & Wakeman, 2013); and variability in student population from year to year (Saven, Anderson, Nese, Farley, & Tindal, 2016). These challenges leave few options for what to report on AA-AAS score reports. Even what may seem like a straightforward status indicator on a large-scale assessment – achievement level in the subject -- may be seen as confusing or having little meaning when parents struggle to understand how their child who has very little evidence of academic knowledge and skills can be "proficient." In a departure from past practices, Developers of the Dynamic Learning Maps (DLM) Alternate Assessment System applied principles from the research literature in designing, developing, and evaluating diagnostic large-scale score reports. The purpose of this paper is to illustrate this process and how decisions were made along the way.

Context

DLM alternate assessments are annually administered to approximately 90,000 students with the most significant cognitive disabilities in 19 states. Assessments measure alternate content standards, called *Essential Elements*, in grades 3-8 and high school in English language arts, mathematics, and science. Students show their knowledge, skills, and understandings on short assessments, called *testlets*, which consist of 3-8 items measuring one or more Essential Element. Testlets measure the Essential Element at one of five[1] *linkage levels* to provide all students with access to grade-level content. The target level measures the grade-level expectation; there are also three precursor levels and one successor level extending beyond the target.

Item responses are scored using diagnostic modeling to determine whether students demonstrated linkage level mastery (see Chapter 5 of DLM Consortium, 2018). Students are classified as masters or non-masters of each linkage level for each tested Essential Element. Mastery statuses are aggregated to report overall mastery in each conceptual area (analogous to a strand). Overall performance in the subject is calculated by applying cuts from a standard setting process (Clark, Nash, Karvonen, & Kingston, 2017) to the total number of linkage levels mastered in the subject. Unlike many large-scale assessments, there is no scale score or concepts commonly reported with a scale score (e.g., standard error of measurement).

DLM assessment results are intended to (a) communicate achievement to a variety of audiences, (b) be included in state accountability models, and (c) be useful and informative to

[1] There are three linkage levels in science: two precursors and the target.

instructional decision making. With the first and third intended uses in mind, we designed individual student score reports to be easy to interpret and use by a variety of audiences.

Individual student score reports summarize results in two parts: a Learning Profile and a Performance Profile. Current operational versions of these profiles are illustrated in Figures 1 and 2, respectively. The Learning Profile summarizes linkage level mastery for every assessed Essential Element. The Performance Profile provides a summary of overall performance in the subject. Conceptual area bar graphs summarize the percent of linkage levels mastered for groups of related Essential Elements, and a performance level summarizes overall achievement in the subject. The Performance Profile also lists the grade- and subject-specific performance level descriptors that indicate the kinds of skills commonly demonstrated by students who score at that performance level. Resources available to support interpretation are described later in this paper.

Given the challenges of past AA-AAS systems and the anticipated unique design of the DLM system, we used an iterative process over a five-year period to develop, refine, and evaluate the reports. The work began in 2012-13 with an understanding that operational results had to be reported starting in 2014-15. We have also conducted studies after the launch of the operational system to evaluate interpretation and uses. The remainder of the paper will describe the process used to develop and evaluate diagnostic score reports. We conclude with future research plans and lessons learned, which may be relevant for other programs interested in pursuing diagnostic score reporting to meet stakeholder needs.

Our Process

This description is grounded in Hambleton and Zenisky's (2013) framework for score report development, which includes seven steps that have also been grouped into three phases:

initial preparation, report development, and report tryout and revision (Zenisky & Hambleton, 2012).

Initial Steps

In Hambleton and Zenisky's (2013) model, the initial steps involve taking in information that can be used to guide the design of original report mockups. These steps include an assessment of key stakeholders' information needs, identification of audiences and how their characteristics as score report consumers inform decisions about information to include in reports, and review of existing score reports that illustrate various design choices and types of information. In the DLM process, we conducted a needs assessment, articulated understandings about audiences, and gathered examples of score reports.

Needs Assessment and Audience Characteristics

During this phase we focused first on parents as the primary stakeholder for individual student score reports. In 2013 DLM staff partnered with staff from The Arc (a national advocacy group for individuals with intellectual and developmental disabilities) to investigate parent experiences with alternate assessments and the special education system (Nitsch, 2013). Forty-four parents of children who took AA-AAS responded to 15 research questions which culminated in over 17 hours of recordings. The research questions followed six domains of inquiry and included parental experience with alternate assessments, academic and nonacademic expectations for their children, postsecondary education goals, and preferred reporting process for information gleaned from AA-AAS.

Analysis of the parent responses to the research questions indicated that parents aimed for their children to attain enough mastery of academic and nonacademic skills to be employable. Parents considered employability focus more favorably than postsecondary education. Many parents viewed academic expectations as less important than development of social or functional skills and thus regarded alternate assessments negatively.

Parents also reported limited awareness and understanding of alternate assessments and challenges with receiving information about alternate assessments from schools. They perceived AA-AAS as a mandate driven by accountability and irrelevant for their child. They indicated that the schools and state, not themselves, were the primary audience for AA-AAS results. Participants noted that they did not receive information about AA-AAS from the school, beyond the IEP team conversation in which the decision was made for the student to participate in AA-AAS instead of the general assessment. Results were reportedly not useful for informing transition planning and were not typically used to inform IEP goals. Some parents' opinions about the AA-AAS were influenced by conversations with teachers, who perceived the assessments as irrelevant, a waste of time, and more of a paperwork burden for teachers.

Parent perceptions of AA-AAS results were mixed. Some reported that they had not received results at all. Others did, but struggled to interpret the results. For example:

> *There's no information about how they got to the scores, so if my son was advanced proficient, advanced proficient in what? ...Scoring-wise, there was no detail as to how they came to the fact that this is what he got, this is how they gauged him, this is what his percent was.* (Nitsch, 2013, p. 23)

Other criticisms of current AA-AAS reports were that the results were not accurate, the terminology was confusing, the messages focused on students' deficits rather than achievements, the results were useless for educational planning, and that reports and supporting materials were not available in languages other than English.

Although parents viewed assessment results with skepticism, they provided helpful recommendations for the future including ways of communicating about assessment results (i.e., paper and online versions, meetings with teachers to discuss the results), additional resources that could support parents' understanding of the contents, and materials that would help them take action at home on goals the student was not mastering.

Example Score Reports

In 2013 the DLM partner states were still administering their state-specific AA-AAS, most of which were based on portfolios or performance tasks and tended to yield information only about overall performance in a subject. While none of their systems were designed like DLM would be, we gathered from state websites examples of their current AA-AAS score reports. We reviewed sample reports for ideas about terminology that might aid interpretation and prevent misinterpretation, for example, the way states interpreted performance relative to extended grade-level content standards – a topic that parents reportedly found confusing. We also sought but could not find published examples of reports on diagnostic assessment results beyond what appeared in previously identified empirical literature.

Design and Development

This phase involves the design of score reports, incorporating information from the previous phase with input from a range of experts (Hambleton & Zenisky, 2013). For the DLM project, the initial phase yielded rich information about parents' information needs and, via the parent focus groups, insight into how educators were communicating with parents about AA-AAS. The review of existing score reports yielded less useful information. We started this phase with some guiding principles for report contents, including:

- Maximizing relevance

- Emphasizing strengths rather than deficits
- Representing complex and unfamiliar concepts and vocabulary in easily interpretable ways

While some questions in the score report review sheet provided by Hambleton & Zenisky (2013) were relevant and useful, some of their questions were based on assumptions that reports were describing scale scores. Review questions that were relevant to the DLM system design were considered during our process.

We treated this phase as highly iterative, with stakeholder feedback at key points. Prototypes were originally designed by DLM staff with expertise in large-scale state assessment implementation, accountability, psychometrics, special education, academic content, cognition, and graphic design. We reviewed early drafts with the DLM Consortium Governance Board, comprised of state education agency staff in assessment and special education, and with the DLM Technical Advisory Committee, comprised of national experts in large-scale assessment, psychometrics, and the student population, in early 2014.

Evaluation

After several drafts, we had prototypes we deemed were ready for input from parents and educators. These versions included early concepts for a Performance Profile and a Learning Profile, each with information about status and growth. Two full prototype sets were provided for the 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. One used shading to indicate all linkage levels mastered, while the other also showed text indicating the skill associated with the highest linkage level mastered and the skill expected at the target level.

Staff from the DLM project and The Arc conducted three parent focus groups to gather feedback on mock score reports (Nitsch, 2014). Parents responded positively to the reports, indicating they were more helpful than their states' current AA-AAS reports. Parents also reacted positively to the norm-referenced information, although they needed clarification on whether the comparison was to all students who took DLM assessments or just those who were similar to their children. Parents understood the contents of Performance Profiles 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. In response to the section showing results by conceptual area, parents appreciated the focus on the student's strengths (what was mastered) rather than deficits (what was not mastered). A section that presented mastery and growth information together, with both graphs and numbers for each, introduced some confusion.

Despite initial efforts to minimize confusing vocabulary, parents found the use of "targets" to be confusing and subjective and suggested providing more information about targets in supplemental materials. Parents also suggested adding definitions and examples to make the language more accessible. To provide more information to those who want it, parents suggested placing a direct link on the score reports to supplemental reports and splitting reports into summaries and more detailed information.

Local educators, including teachers and district-level staff, also provided feedback on the mock Performance Profiles and Learning Profiles. Conceptual-area results were viewed as holding strong potential to be useful when describing students' present levels of performance on IEPs. Educators objected to including normative information on the reports, expressing concern about parents whose children would not perform well in comparison to their peers.

Focus group feedback informed a final version of prototype reports that were used to check interpretability among educators. Five current teachers were given a brief overview of DLM system design and terminology, then were given a sample Learning Profile and Performance Profile. After being given a few minutes to independently review the reports, they had an opportunity to ask clarifying 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. Participants tended to focus on text-based descriptions of academics, not quantitative information or the performance level descriptors. They primarily discussed the linkage level descriptors on the Learning Profile, leading with students' areas of mastery and then describing higher, un-mastered levels as areas to focus on in the future. Some also expressed broad strengths and areas for growth based on the conceptual-area results.

Initial Operational Version

Work at this phase culminated in the initial version of operational score reports delivered in 2014-15. Notable changes driven by stakeholder feedback included: (a) simplifying some terminology (e.g., replacing "levels" with "skills"); (b) expanding introductory text to support understanding of how Essential Elements relate to grade-level standards and including caveats to prevent likely misinterpretation of what the results represented; (c) adding text descriptors to every linkage level for every Essential Element on the Learning Profile; and (d) removing normative comparisons, which were deemed a significant source of risk of misinterpretation and counter to the goal of reporting student competencies rather than deficits.

We also prepared several supporting materials that states could tailor and use as desired. We created a score report interpretation guide with screen shots and callouts to support understanding of key parts of the report. We also prepared a cover letter that could accompany the DLM score report and be signed by superintendents, and guides for state and local education leaders that succinctly explained the reports along with intended interpretations and cautions against likely misinterpretations. The DLM Governance Board reviewed and provided feedback on drafts of all of these materials.

Implementation

Hambleton and Zenisky (2013) encourage systematic field testing to evaluate reports and use results to improve future versions. They recommended stakeholder focus groups to evaluate usability and comprehension, and experimental designs to evaluate how contrasting examples of report contents and format impact understanding.

Given timeline and resource constraints, and the intended uses of DLM results compared with past AA-AAS, we approached this phase a bit differently. We prioritized a series of studies that primarily involved teachers, for multiple reasons:

- Earlier needs assessments and parent focus group confirmed our impression about how much parents rely on teachers to provide access to and shape interpretation of AA-AAS results
- Instructional use is a key goal in the DLM system, and teachers ultimately control the extent to which this goal is met

Also, consistent with the argument made by O’Leary and colleagues (2017), we put interpretation and use at the forefront of the studies, with understanding of report contents as a secondary goal that could be accomplished while observing interpretations.

In 2016 and 2017 we conducted a study to examine how teachers interpreted and used DLM individual summative score reports (comprised of the Performance Profile and Learning Profile). The study explored teacher understanding of the information presented in the reports, teacher explanations about the reports to parents, teacher resources helpful in score report use, and teacher use of these reports in conjunction with educational planning (Karvonen et al., 2016).

Participants included 12 teachers from two states and two parent advocates from one state. Data sources were individual and paired interviews. Individual interviews focused on interpretation and explanation of the results to parents. Paired interviews focused on interpretation and use of the results to plan for instruction. Both types of interviews used operational versions of score reports with fictitious data showing contrasting patterns of achievement. Specific materials were tailored to each participant so they viewed reports for the grades and subjects they currently taught. Participants were encouraged to think aloud as they interpreted the reports. Probing questions were used to encourage verbalization and to clarify responses. Transcripts were coded for research question addressed, part of report the participant was referring to, and thematic codes related to each research question.

Overall, teachers found the reports helpful for instructional planning and communicating with parents. It was difficult to verify their understanding of the concepts because of their “reliance on the exact text in the report, preference for mastery statements over aggregated information, and a tendency to not rephrase key meanings” (Karvonen et al., 2016, pp. 16).

Researchers noted some misconceptions including the distinction between percentile and percent and between performance level labels and linkage level labels. Most participants used the mastery list in the score report to explain specific skills to parents and used the learning profile to describe next steps for instruction (Karvonen et al., 2017). One potential pitfall identified for future examination was teachers' understanding of how progressions of skills related to the current grade's Essential Elements related to instruction they would begin in the next grade, when the student would be working toward different grade-level Essential Elements.

In 2017 we also evaluated the impact of resources intended to improve teacher understanding of the score reports. These resources included a PDF guide and a video tutorial. To evaluate the value of the resource, the teachers were given a pre and posttest on score report interpretation. They also had the opportunity to retake the quiz to achieve a passing score. Ninety-three teachers from across states began the study, but there was substantial attrition. After the viewing the resources, 39% of completers passed the posttest. Furthermore, 41% of teachers who rated themselves as neutral or negative regarding their confidence in interpreting score reports prior to the tutorial passed the posttest. Analyses of percent of correct responses to each item helped staff identify priority areas for enhancements to score report design, tutorial, and additional resources.

In spring 2018 we conducted a series of individual and small focus group interviews to evaluate teachers' use of DLM score reports. This study differed from 2016 and 2017, which were based on interpretation of fictitious reports. In 2018 we asked about actual uses. These results are summarized in the first (Clark & Kingston, 2019) and fourth (Nehler, Clark, & Burnes, 2019) papers in this symposium.

Discussion

Treating this series of studies as a coherent set, with a long-term plan that evolves as we learned lessons from each study, has allowed us to regularly probe for potential threats to validity related to stakeholder interpretation and use of results. The early prototypes underwent quite a few changes before the reports became operational. The operational versions have undergone minor changes, such as updated footer text to aid interpretation and point readers to a website with resources, and changing Learning Profile shading to distinguish untested versus un-mastered linkage levels. We also revised the score report training to be four separate modules teachers can access based on their specific information needs, with contents revised based on misconceptions captured in the posttest responses. We found the Hambleton & Zenisky (2013) framework helpful in building the long-term design and evaluation plan, and agree with O'Leary et al. (2017) about the criticality of systematically evaluating actual interpretation and use. Our work has also benefitted from built-in feedback mechanisms including review of study plans and results by the DLM Governance Board and the DLM Technical Advisory Committee.

The initial steps and design/development stage brought both challenges and opportunities. Needs assessment focus groups revealed entrenched feelings of lack of relevance of AA-AAS for teachers, students, and parents. The fact that the DLM system would yield diagnostic information that described what students could do instead of their deficits was one step in bridging that gap. Because none of our partner states' previous AA-AAS yielded unidimensional scores, we did not have to overcome a history about expected information such as scale score and measurement error. However, the teacher interviews conducted in 2015-2017 uncovered common misconceptions associated with special education – ones that would not have appeared in the published literature. These included confusing percent of skills mastered with common instructional percentages around the use of massed trials in instruction or the wording of IEP goals based on frequency and accuracy of demonstrated skills. Besides redesigning training to address these misconceptions, we will monitor to see if some of those misconceptions fade once the DLM assessment system is well-established and the score reports are more familiar.

To the question of whether teachers interpret and use report contents as intended for instructional purposes, we have mixed findings so far. Teachers demonstrate an understanding of how to use the reports, and use different parts for different purposes. Their interpretations were generally correct. Yet they rely on the text itself and do not tend to paraphrase. If they rely on academic skill statements and can interpret those correctly in generalities, how much do misinterpretations of precise language matter for their instructional planning? One of the potential pitfalls in instructional use is the risk of unintended consequences: if teachers view each linkage level statement as a discrete skill, they may continue providing instruction using that model (which has been dominant for decades in instruction for this population) rather than shift to the more coherent, conceptually-focused instruction intended for standards-aligned instruction in the DLM system.

So far our findings on operational use are largely confirmatory, and based on teachers who are likely to be optimal users of the score reports. We don't yet know about score report interpretations and uses among teachers who would not necessarily volunteer for studies. Our ability to gather data from a more representative sample on actual instructional uses is hindered when teachers do not actually receive their students' score reports (Clark, Karvonen, Swinburne Romine, & Kingston, 2018). We face similar challenges as other score report developers in gaining access to representative and diverse populations. We did include parent advocates and non-English speaking parents in hopes of meeting those audiences' needs, but we do not presume the current reports fully meet the information needs of all parents. Further investigation is also needed on the usefulness of ancillary interpretive materials for various audiences, and on interpretation and use of aggregated reports. If we eventually find certain report elements cause misinterpretations or faulty inferences, we could then use a more experimental approach to evaluate whether a different design improves accurate interpretations and inferences.

The DLM assessment is the first K-12 large-scale assessment for accountability that has results based on mastery classifications. We are still working on effective ways to communicate complex concepts, such as measurement uncertainty. Tension also remains between the desire to add more information to aid with interpretation and ensuring reports are succinct enough that audiences will use them. As we shift into online reports, including dynamic user-driven displays (see Dolan et al., this symposium), we have an opportunity to link to more information for those who want to know more. This transition will require additional evaluations in which we gather evidence of users' information processing and system interactions along with their interpretations and uses of report contents. This type of direction also provides opportunities to contribute to the field, for example by updating Hambleton and Zenisky (2013) with criteria for diagnostic reports and processes for evaluating online, dynamic reports.

References

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

Clark, A. K., Karvonen, M., Kingston, N., Anderson, G., & Wells-Moreaux, S. (2015, April). *Designing alternate assessment score reports that maximize instructional impact*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, Illinois.

Clark, A. K., Karvonen, M., Swinburne Romine, R., & Kingston, N. M. (2018, April). *Teacher use of score reports for instructional decision-making: Preliminary findings.* Paper presented at the annual meeting of the National Council on Measurement in Education, New York, NY.

Clark, A. K., & Kingston, N. M. (2019, April). *Diagnostic assessment results: Instructional uses and potential pitfalls*. Paper presented at the annual meeting of the American Educational Research Association, Toronto, ON.

Clark, A. K., Nash, B., Karvonen, M., & Kingston, N. (2017). Condensed mastery profile method for setting standards for diagnostic assessment systems. *Educational Measurement: Issues and Practice, 36*(4), 5-15. doi: 10.1111/emip.12162

Dynamic Learning Maps Consortium. (2018). *2017-2018 technical manual update – Integrated model.* Lawrence, KS: University of Kansas, ATLAS. Retrieved from https://dynamiclearningmaps.org/about/research/publications

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,* 145-220.

Hambleton, R. K., & Zenisky, A. L. (2013). Reporting test scores in more meaningful ways: A research-based approach to score report design. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J. C. Hansen, N. R. Kuncel, S. P. Reise, & M. C. Rodriguez (Eds.), *APA handbooks in psychology. APA handbook of testing and assessment in psychology, Vol. 3. Testing and assessment in school psychology and education* (pp. 479-494). Washington, DC, US: American Psychological Association.

Hoover, N. R., & Abrams, L. M. (2013). Teachers' instructional use of summative student assessment data. *Applied Measurement in Education, 26,* 219-231

Karvonen, M., Clark, A. K., & Kingston, N., (2016, April). *Alternate assessment score report interpretation and use: Implications for instructional planning*. Presentation at the 2016 annual meeting of the National Council on Measurement in Education, Washington, DC.

Karvonen, M., Swinburne Romine, R., Clark, A. K., Brussow, J., & Kingston, N. (2017, April). *Promoting accurate score report interpretation*. Presentation at the 2017 annual meeting of the National Council on Measurement in Education, San Antonio, TX.

Kingston, N. M., Karvonen, M., Bechard, S., & Erickson, K. (2016). The philosophical underpinnings and key features of the Dynamic Learning Maps Alternate Assessment. *Teachers College Record (Yearbook), 118*(14). Retrieved from http://www.tcrecord.org

Leighton, J. P., & Gierl, M. J. (Eds.). (2007). *Cognitive diagnostic assessment for education: Theory and applications*. New York, NY: Cambridge University Press.

Marion, S. F. (2018). The opportunities and challenges of a systems approach to assessment. *Educational Measurement: Issues and Practice, 37*(1), 45–48.

Nehler, C., Clark, A. K., Burnes, J. (2019, April). *Evaluation of resources to support diagnostic score report interpretation.* Paper presented at the annual meeting of the American Educational Research Association, Toronto, ON.

O'Leary, T. M., Hattie, J. A. C., & Griffin, P. (2017). Actual interpretations and use of scores as aspects of validity. *Educational Measurement: Issues and Practice, 36*(2), 16-23. doi: 10.1111/emip.12141

Nitsch, C. (2013). *Dynamic Learning Maps: The Arc parent focus groups*. Unpublished manuscript. Washington, DC: The Arc. Retrieved from https://dynamiclearningmaps.org/about/research/publications

Nitsch, C. (2014, September 14). *Parent focus groups: Feedback on mock score reports.* Unpublished manuscript. Washington, DC: The Arc. Retrieved from https://dynamiclearningmaps.org/about/research/publications

Roberts M. R., & Gierl, M. J. (2010). Developing score reports for cognitive diagnostic assessments. *Educational Measurement: Issues and Practice, 29*(3), 25-38. https://doi.org/10.1111/j.1745-3992.2010.00181.x

Rupp, A., Templin, J., & Henson, R. (2010). *Diagnostic measurement: Theory, methods, and applications.* New York, NY: The Guilford Press.

Sinharay, S., & Haberman, S. J. (2009). How much can we reliably know about what examinees know? *Measurement:Interdisciplinary Research and Perspectives, 7 7*, 46-49. doi: 10.1080/15366360802715486

Sinharay, S., Puhan, G., & Haberman, S. J. (2010). Reporting diagnostic scores in educational testing: Temptations, pitfalls, and some solutions. *Multivariate Behavioral Research, 45*(3), 553-573. doi: 10.1080/00273171.2010.483382

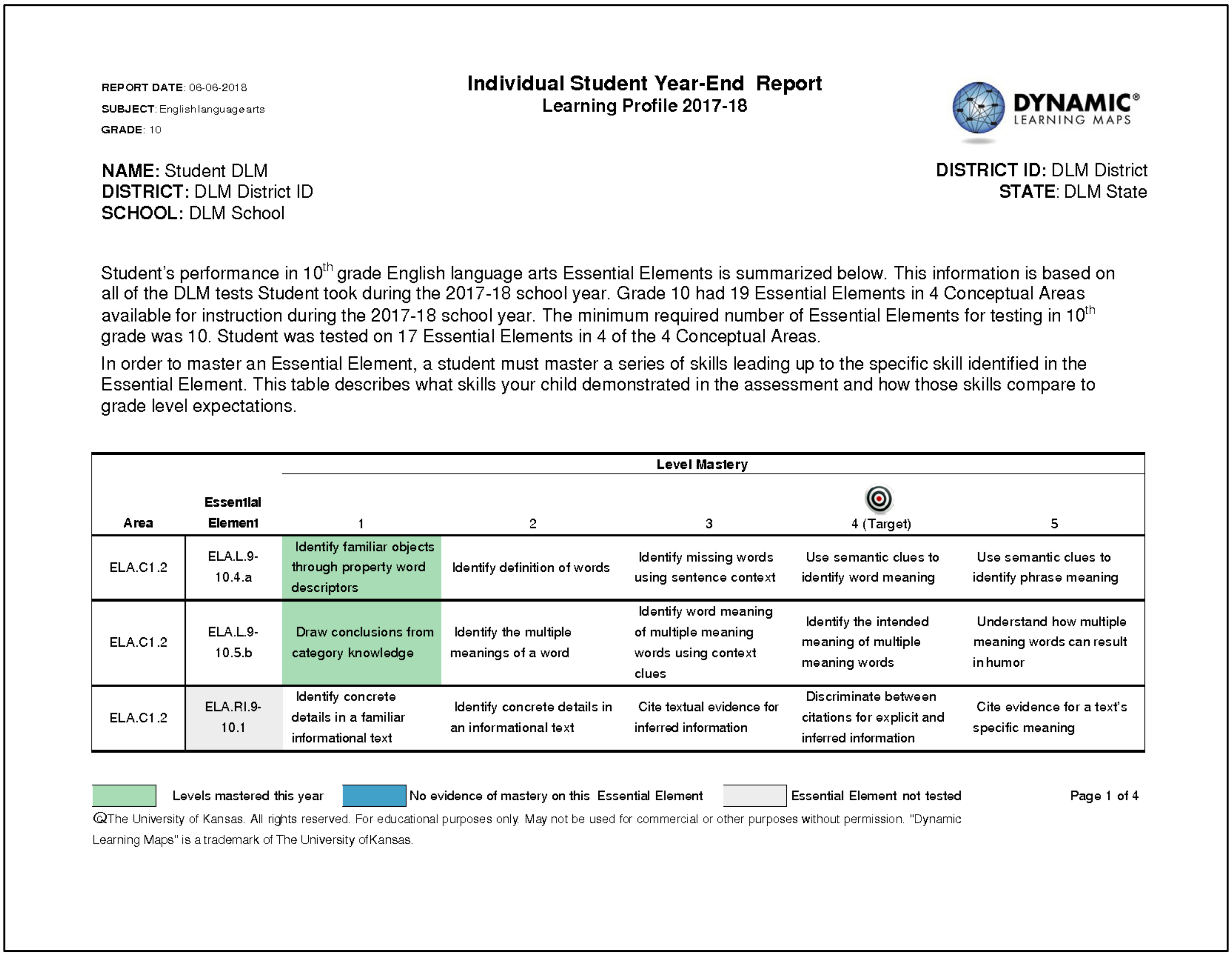
Saven, J. L., Anderson, D., Nese, J. F., Farley, D., & Tindal, G. (2016). Patterns of statewide test participation for students with significant cognitive disabilities. *The Journal of Special Education, 49*(4), 209–220. doi:10.1177/0022466915582213

Wilson, M. (2018). Making measurement important for education: The crucial role of classroom assessment. *Educational Measurement: Issues and Practice, 37*(1), 5–20.

Yeh, S. S. (2006). Reforming federal testing policy to support teaching and learning. *Educational Policy, 20,* 495-524.

Zapata-Rivera, J. D., & Katz, I. R. (2014). Keeping your audience in mind: Applying audience analysis to the design of interactive score reports. *Assessment in Education: Principles, Policy & Practice, 21*, 442-463. doi: 10.1080/0969594X.2014.936357

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.



REPORT DATE: 06-06-2018
SUBJECT: English language arts
GRADE: 10

Individual Student Year-End Report
Learning Profile 2017-18

DYNAMIC®
LEARNING MAPS

NAME: Student DLM
DISTRICT: DLM District ID
SCHOOL: DLM School

DISTRICT ID: DLM District
STATE: DLM State

Student's performance in 10th grade English language arts Essential Elements is summarized below. This information is based on all of the DLM tests Student took during the 2017-18 school year. Grade 10 had 19 Essential Elements in 4 Conceptual Areas available for instruction during the 2017-18 school year. The minimum required number of Essential Elements for testing in 10th grade was 10. Student was tested on 17 Essential Elements in 4 of the 4 Conceptual Areas.

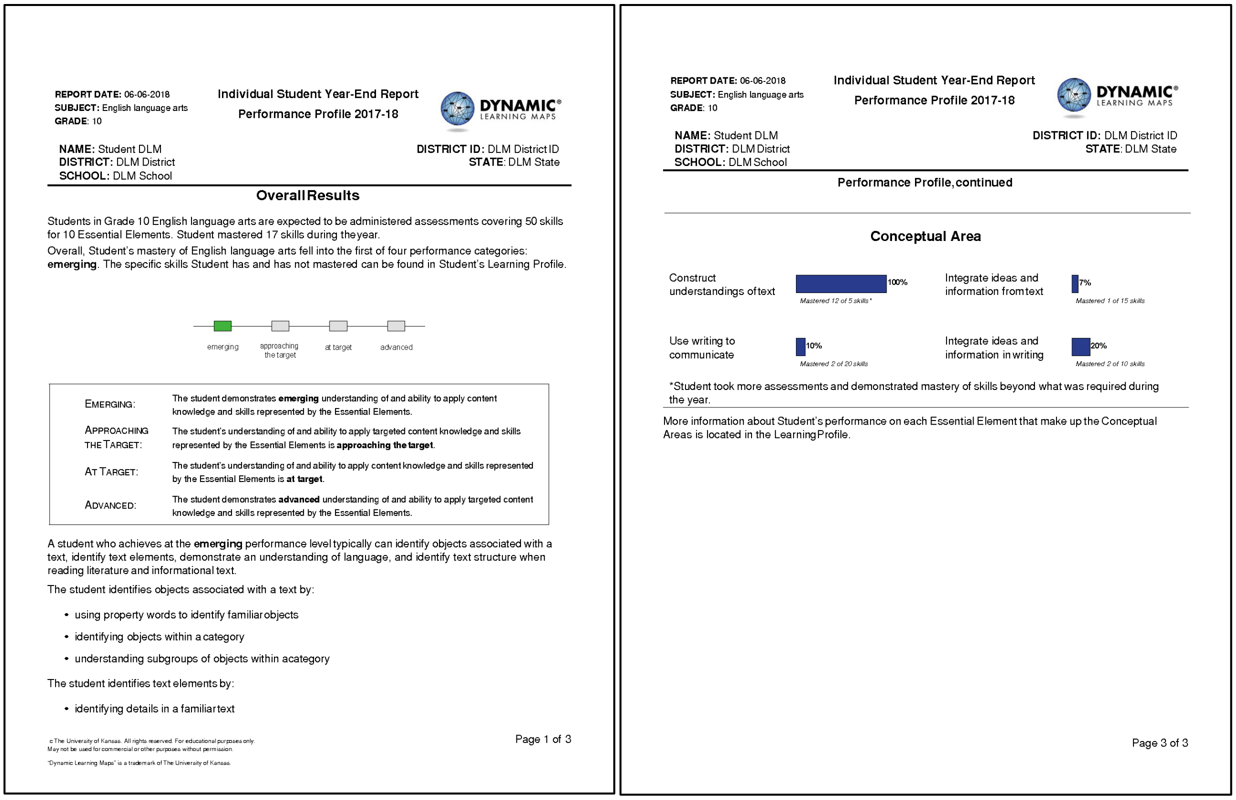
In order to master an Essential Element, a student must master a series of skills leading up to the specific skill identified in the Essential Element. This table describes what skills your child demonstrated in the assessment and how those skills compare to grade level expectations.

		Level Mastery				
Area	**Essential Element**	1	2	3	◎ 4 (Target)	5
ELA.C1.2	ELA.L.9-10.4.a	Identify familiar objects through property word descriptors	Identify definition of words	Identify missing words using sentence context	Use semantic clues to identify word meaning	Use semantic clues to identify phrase meaning
ELA.C1.2	ELA.L.9-10.5.b	Draw conclusions from category knowledge	Identify the multiple meanings of a word	Identify word meaning of multiple meaning words using context clues	Identify the intended meaning of multiple meaning words	Understand how multiple meaning words can result in humor
ELA.C1.2	ELA.RI.9-10.1	Identify concrete details in a familiar informational text	Identify concrete details in an informational text	Cite textual evidence for inferred information	Discriminate between citations for explicit and inferred information	Cite evidence for a text's specific meaning

🟩 Levels mastered this year 🟦 No evidence of mastery on this Essential Element ⬜ Essential Element not tested Page 1 of 4

©The University of Kansas. All rights reserved. For educational purposes only. May not be used for commercial or other purposes without permission. "Dynamic Learning Maps" is a trademark of The University of Kansas.

Figure 1. The Learning Profile portion of individual student score reports indicates linkage levels mastered for the five complexity levels for each Essential Element.



REPORT DATE: 06-06-2018
SUBJECT: English language arts
GRADE: 10

Individual Student Year-End Report
Performance Profile 2017-18

DYNAMIC® LEARNING MAPS

NAME: Student DLM
DISTRICT: DLM District
SCHOOL: DLM School

DISTRICT ID: DLM District ID
STATE: DLM State

Overall Results

Students in Grade 10 English language arts are expected to be administered assessments covering 50 skills for 10 Essential Elements. Student mastered 17 skills during the year.

Overall, Student's mastery of English language arts fell into the first of four performance categories: **emerging**. The specific skills Student has and has not mastered can be found in Student's Learning Profile.

EMERGING:	The student demonstrates **emerging** understanding of and ability to apply content knowledge and skills represented by the Essential Elements.
APPROACHING THE 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.

A student who achieves at the **emerging** performance level typically can identify objects associated with a text, identify text elements, demonstrate an understanding of language, and identify text structure when reading literature and informational text.

The student identifies objects associated with a text by:

- using property words to identify familiar objects
- identifying objects within a category
- understanding subgroups of objects within acategory

The student identifies text elements by:

- identifying details in a familiartext

Page 1 of 3

REPORT DATE: 06-06-2018
SUBJECT: English language arts
GRADE: 10

Individual Student Year-End Report
Performance Profile 2017-18

DYNAMIC® LEARNING MAPS

NAME: Student DLM
DISTRICT: DLM District
SCHOOL: DLM School

DISTRICT ID: DLM District ID
STATE: DLM State

Performance Profile, continued

Conceptual Area

Construct understandings of text
100%
*Mastered 12 of 5 skills**
Integrate ideas and information from text
7%
Mastered 1 of 15 skills
Use writing to communicate
10%
Mastered 2 of 20 skills
Integrate ideas and information in writing
20%
Mastered 2 of 10 skills

*Student took more assessments and demonstrated mastery of skills beyond what was required during the year.

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

Page 3 of 3

Figure 2. The Performance Profile portion of individual student score reports includes the performance level for the subject, performance level descriptors describing skills typical of students achieving at that level, and conceptual area bar graphs summarizing the percent of skills mastered in each area of related standards.