Diagnostic Assessment Results: Instructional Uses and Potential Pitfalls

Amy K. Clark

Neal M. Kingston

University of Kansas

Author Note

Paper presented at the 2019 annual meeting of the American Educational Research Association, Toronto, ON. Correspondence concerning this paper should be addressed to Amy Clark (akclark@ku.edu).

Abstract

Stakeholders increasingly call for assessment results that are useful to instructional practice. While performance levels and scale score values have served accountability purposes well, they are often at too large a grain size to be instructionally useful. Diagnostic assessments have emerged as a measurement approach that can provide more detailed results to stakeholders. This paper highlights important considerations related to diagnostic assessment results and their potential for instructional use, as well as pitfalls around interpretation that should be avoided or addressed by diagnostic assessment systems. We provide context using the Dynamic Learning Maps Alternate Assessment System, a large-scale diagnostic assessment from which results are used for both instructional and accountability reporting purposes. We also provide discussion for applying lessons learned from this assessment to other applications of diagnostic assessments.

Keywords: diagnostic assessment, score reporting, interpretation and use, instructional use

Diagnostic Assessment Results: Instructional Uses and Potential Pitfalls

Assessments based on psychometric diagnostic classification models are a valuable tool for student learning and educator instructional practice. One key benefit is that such assessments can be built to provide fine-grained information about specific attributes, or skills, measured by the assessment. Despite their recent focus in measurement literature and research-based applications, there are limited operational diagnostic assessments currently in use in large-scale academic assessment contexts. This paper introduces diagnostic assessment systems, including their unique scoring and reporting considerations, in light of interpretations and uses of results and guidance provided in professional standards (American Educational Research Association [AERA], American Psychological Association, & National Council on Measurement in Education, 2014). We discuss the context for an operational, large-scale diagnostic assessment system used in 18 states, the Dynamic Learning Maps Alternate Assessment System. Further, we discuss implications for other diagnostic assessment contexts, including instructional uses and potential pitfalls related to reporting results based on probabilities of mastery and overall profiles of mastery rather than raw or scale score values.

Using Assessment Results

Since the passage of the No Child Left Behind Act in 2002, states have been required to administer educational assessments for accountability purposes. In addition to inclusion in accountability models, states and districts often use aggregated results from summative assessments for program evaluation and resource allocation. While performance-level results from traditional large-scale assessments serve these purposes well, one of their historic limitations is that overall achievement results have limited utility for informing educators' instructional practice. The grain size of reported results is often too coarse to be instructionally

useful (Marion, 2018). Furthermore, the reporting timeline is often such that reports are delivered in the following academic year, after students have advanced a grade, are being taught new grade-level standards, and thus provide no benefit to the learning of those students.

Recent measurement and technological advancements have produced assessments that are increasingly student-centered, including diagnostic assessments and computer adaptive measures. Stakeholders increasingly desire reports that are similarly nuanced and can be used to inform instructional planning, monitoring, and adjustment.

Use of Large-Scale Assessment Results for Instruction

There is a limited body of research on how educators use assessment results to inform instruction. Yeh (2006) found that 56 of 61 interviewees (92%) were concerned that score reports from a federally-mandated assessment provided inadequate diagnostic information about student knowledge, skills, and understandings. Interviews also noted that because results were from the prior academic year, they were less informative to the current year's instruction. A similar study surveying teachers on their use of summative score reports found that teachers most frequently evaluated aggregated student results by examining the mean or mode, and less frequently disaggregated results for student subgroups or by content standard (Hoover & Abrams, 2013). These findings indicate that teachers did not use results in ways that would likely provide strong support for instructional practice, such as informing specific plans for instruction regarding student intervention or enrichment, or planning instructional groupings based on areas to target instruction across students.

Diagnostic Assessments

A key benefit to administering diagnostic assessments is that results provide fine-grained information about student mastery. Rather than reporting an overall performance level or a single

raw or scale score value, diagnostic assessments produce mastery profiles that summarize specific skill mastery information. Mastery is determined from probabilistic scoring models that determine the likelihood that students mastered each skill measured by the assessment (e.g., diagnostic classification modeling; Bradshaw, 2017; Rupp et al., 2012). Probability values closer to 1.0 indicate greater certainty that the student mastered the skill, while values near 0.5 indicate the model cannot discern whether the student demonstrated mastery of the skill from their item response pattern. Often these probability values are replaced with dichotomous mastery determinations for each skill. This process requires a threshold or cut to distinguish masters from nonmasters. For example, setting a cut at 0.8 indicates that all students achieving a mastery probability of that magnitude or greater are considered masters of the skill, while anyone with a mastery probability below that value is considered a nonmaster.

Reporting for diagnostic assessments is typically provided via a mastery profile. The report shows each attribute or skill measured by the assessment. Reports may indicate the probability value and/or whether the student achieved mastery (e.g., Rupp et al., 2012, pp. 66; Bradshaw, 2017, pp. 316). However, because of their difference from traditional reporting methods focusing on overall performance in the subject, diagnostic assessment results may also be prone to misunderstanding or misinterpretation. Since one intended purposes of diagnostic assessment results is that they are useful for instruction, the *Standards for Educational and Psychological Testing* (AERA et al., 2014) indicate that evidence should be collected to evaluate the extent that results are used as intended.

Dynamic Learning Maps Assessments

The Dynamic Learning Maps Consortium administers diagnostic alternate assessments to students with the most significant cognitive disabilities in 18 states. Alternate content standards

are available at five complexity levels to provide all students access to grade-level content. Assessments are scored using latent class models to determine the probability that students have mastered each assessed skill (see Chapter 5 of DLM Consortium, 2018 for scoring method). The basis of reporting is the set of skills mastered across all grade-level content standards. Results are summarized as both (1) fine-grained skill mastery decisions (mastered or not-mastered; Figure 1), which are provided in both within-year progress reports and in summative score reports, and (2) aggregated skill mastery information across sets of conceptually-related standards and in the subject overall (Figure 2), which are only provided in summative score reports. Cuts between the total skills mastered delineate overall performance levels (Clark, Nash, Karvonen, & Kingston, 2017), which are used in state accountability models. Prior to administering assessments, teachers complete required training, which briefly introduces DLM scoring and mastery determinations, among other concepts.

DLM diagnostic score reports were designed after multiple points of feedback from parents (Nitsch, 2013) and educators (Clark, Karvonen, Kingston, Anderson, & Wells-Moreuax, 2015; Karvonen, Clark, & Kingston, 2016; Karvonen, Swinburne Romine, Clark, Brussow, & Kingston, 2017) and after consulting best practice in the literature (e.g., Hambleton & Zenisky, 2012). More detail on the score report design and evaluation process is provided in paper #2 of this session (Karvonen, Clark, Swinburne Romine, & Kingston, 2019).

Data Collection

During spring 2018, we recruited teachers who had administered DLM assessments and received summative score reports for the prior academic year to participate in focus groups. Eight focus groups were conducted with a total of 17 teachers from three consortium states.

Sessions focused on interpretation and use of the prior year's summative results in the subsequent academic year.

Instructional Uses

Teachers described varying levels of utility of summative diagnostic reports for planning instruction in the subsequent academic year, with differences observed by grade level. Our sample of teachers of elementary and middle school students whose accountability requirements included annual assessment reported diagnostic reports to be more useful for informing instruction than high school teachers, where students are typically only required to assess in a single grade for state accountability purposes (e.g., 11th grade). Teachers noted challenges when the most recent summative score report available was from several grades prior, particularly for their 11th grade students who only had 8th grade reports available. Teachers also pointed out that often the curriculum in 12th grade, as students taking DLM assessments prepare to transition to college, career, and community participation, was markedly different from the 11th grade curriculum, and therefore results from the prior year were not as useful to instruction. In contrast, elementary and middle school teachers, reported much more utility in using the fine-grained diagnostic reports for instructional decision-making, including specifying IEP goals and planning instructional groupings.

Instructional Planning. Teachers in the focus groups described their processes for using fine-grained summative diagnostic assessment results to create instructional plans in the subsequent academic year. They described evaluating the skills mastered in the prior grade and comparing those to skills available in the current grade's content standards. Prioritization of specific skills for instruction varied based on individual student needs. For some students, teachers described focusing less on skills that the student had already mastered in the prior year

to provide greater breadth of instruction and assessment in the subject; for others, they described prioritizing the next level of skill acquisition within a similar content standard to provide greater depth of instruction and assessment in the subject.

Instructional Groupings. Teachers described the benefit of using diagnostic results to plan instructional groupings. Teachers mentioned using fine-grained skill mastery information on summative reports to plan instruction for students working on the same skills, both within and across standards. For example, teachers described looking at the levels of mastery for each standard and determining groups of students working at the same level. They also looked across standards to identify areas where small group instruction could be effective (e.g., combining across reading for information and reading literature mastery information when the skills measured were similar).

IEP Goals. During focus groups, teachers described using diagnostic score reports to inform IEP goal planning. As one teacher stated, “Their IEP goals are very similar to their [skill mastery statement]. I can say, ‘Hey, let’s look at this [skill] and let’s look at this target [grade-level expectation] and this is what we’re working on in your IEP.’ It’s real easy for me to tie all these things together so we don’t have this weird zigzag of skills. [It’s] more streamlined and better growth.” She went on to say, “I really feel like this holds kids to a higher standard. I think it keeps teachers from writing copout goals.” In other words, showing the five levels, including the grade-level expectation, for each content standard allowed her to show the path toward grade-level expectation for each individual standard, but also examine mastery across standards to show how skills fit together, rather than just a seemingly random list of goals on the IEP.

Potential Pitfalls

While teachers overall emphasized several different instructional uses for diagnostic score reports, both within and across academic years, the unique measurement model and reporting structure provides areas of potential misinterpretation. Diagnostic assessment results look very different than traditional raw or scale score values. Through rounds of focus groups in the design phase, parents and educators indicated a preference for dichotomous mastery shading over raw probability values. Even with this simplification in place, the breadth of information reported, which for DLM assessments includes five levels of mastery for each content standard, can be a lot to interpret and make meaning of. This simplification also removes information about the certainty of mastery decisions, which is commonly reported in traditional assessment score reports and a recommendation in the *Standards for Educational and Psychological Testing* (AERA et al., 2014).

Teachers sometimes demonstrated concrete misinterpretations during focus group discussions. For instance, sections of diagnostic reports in Figure 2 that describe the percent of skills students mastered in related *conceptual areas* were often interpreted as a percent of items correctly answered or as the percent of trials in which a student successfully demonstrated a skill. Other teachers referred to the so-called "black box" of scoring and not understanding how mastery decisions were determined or how the overall performance level was determined from the collection of mastery information presented. Teachers repeatedly expressed an interest in more training and resources to better support their interpretation and use of results.

While teachers generally found the fine-grained information helpful for individualized instructional planning, they also expressed a desire for additional reports. When discussing instructional grouping of students, teachers described examining skill mastery information across

multiple students, and therefore a number of pages or reports. One teacher expressed a desire for an aggregated report that made instructional groupings clearer, particularly around standards and levels on which students were working in common.

In general, teachers still used the traditional assessment language to describe results in terms of "scores" on the assessment, despite the lack of a scale or total score value on the report. Further, while teachers seemed to demonstrate accurate conceptions of what skill mastery when describing how they use results and think about what students know and can do, they did not question the certainty of mastery decisions reflected in the score report, despite their being based on probability values.

Discussion

With the increasing prominence of diagnostic assessment systems, attention must be given to reporting practices for communicating fine-grained assessment results to a variety of audiences. While diagnostic reports pose great potential for supporting teachers in their instructional practice, developers must also use caution when designing reports to ensure that information about student mastery is conveyed in an interpretable and meaningful way.

The findings in the present study demonstrate the instructional use of diagnostic score reports that provide fine-grained results. Teachers described the utility of results to inform instructional planning, create instructional groupings, and formulate and monitor progress toward IEP goals. Even when summative reports were delivered in the subsequent academic year, teachers described making connections to current grade-level expectations and providing individualized and small-group instruction targeted at ensuring all students made academic progress.

These findings also have important implications for score report resources, such as interpretation guides and manuals that support appropriate interpretation of skill-mastery information presented in score reports. Because of the unique scoring method and grain-size of reporting, diagnostic assessment score-report developers must make sure that interpretation guides and other resources adequately explain the scoring process and contents of reports with just the right level of detail for various audiences (i.e., neither too much nor too little detail). District staff may be unfamiliar with the scoring methods and may experience challenges knowing what resources would be useful to teachers or how to provide adequate training on the reporting method for diagnostic measures. While teachers in this study indicated a desire for more training, time and availability of district staff to provide numerous training opportunities may be limited.

To address these challenges, large-scale assessment systems can make resources available to support assessment administration and interpretation of results with fidelity. This might include readily available materials that districts can point schools toward for use during professional development and professional learning community activities. Districts could further leverage these resources during district-provided in-service training to better equip teachers in the classroom to use results to inform instruction and share pertinent information with teachers during IEP meetings and conferences.

While this study highlights several opportunities for improvements, it also identifies the potential for large-scale diagnostic assessment results to have an important and meaningful impact on students and teachers in the classroom. As Randy Bennett (2018) indicated in his recent National Council on Measurement in Education presidential address, the field of measurement is heading toward assessments that, among other things, attempt to improve student

learning and provide more effective reporting, both of which are areas where diagnostic assessment systems demonstrate great potential.

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.

Bennett, R. (2018). Educational assessment: What to watch in a rapidly changing world. *Educational Measurement: Issues and Practice, 37*(4), 7-15.

Bradshaw, L. (2017). Diagnostic classification models. In A. A. Rupp & J. P. Leighton (Eds.) *The handbook of cognition and assessment: Frameworks, methodologies, and applications*, pp. 297-327. Malden, MA: Wiley.

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., 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.

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

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*. Paper presented at the annual meeting of the National Council on Measurement in Education, Washington, DC.

Karvonen, M., Clark, A. K., Swinburne Romine, R., & Kingston, N. (2019, April). *Development and evaluation of diagnostic score reports: Process and recommendations.* Paper presented at the annual meeting of the American Educational Research Association, Toronto, ON.

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

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

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

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

No Child Left Behind Act of 2001, P.L. 107-110, 20 U.S.C. § 6301 et seq. (2002).

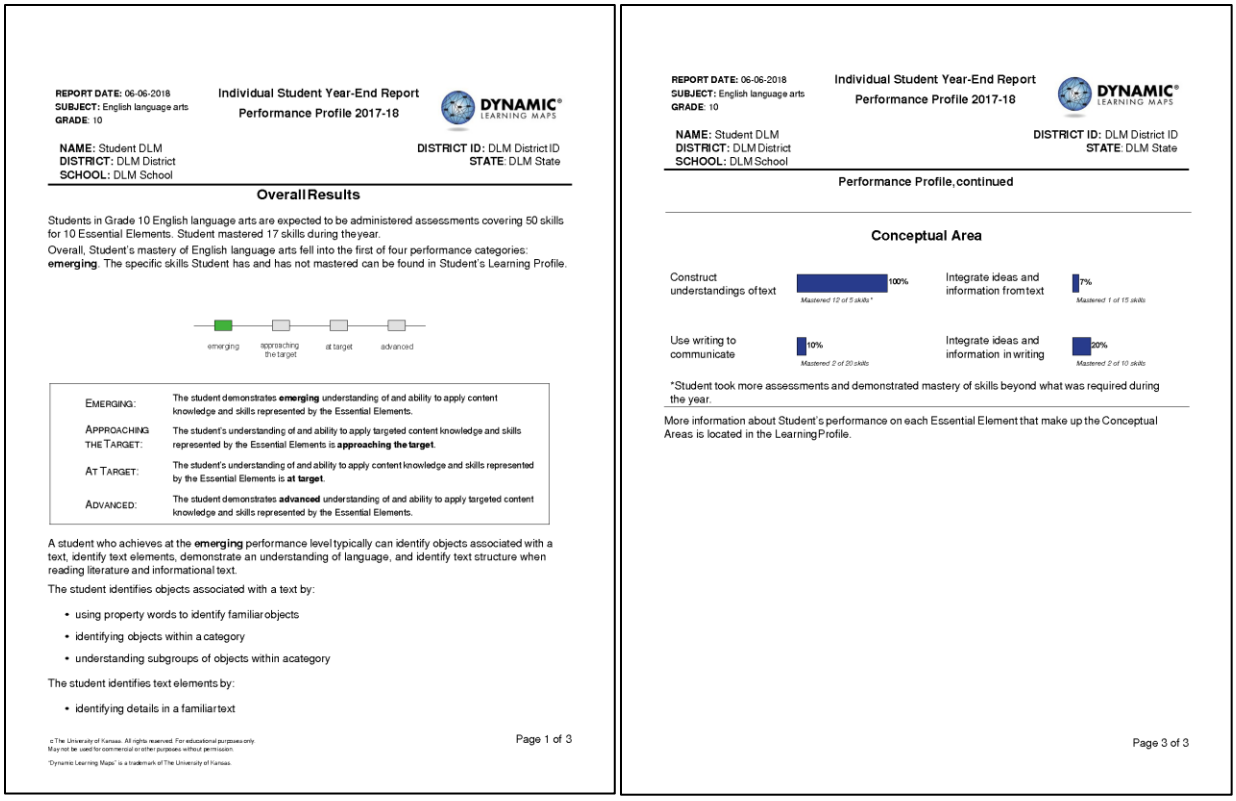
Rupp, A., Templin, J., & Henson, R. (2010). *Diagnostic measurement: Theory, methods, and applications.* New York, NY: The Guilford Press.

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

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.

Area	Essential Element	Level Mastery				
		1	2	3	4 (Target)	5
ELA.C1.2	ELA.RI.9-10.2	Identify concrete details in an informational text	Identify details relevant to the topic of text	Summarize a familiar informative text	Identify key details supporting the central idea	Support implicit and explicit meaning with details
ELA.C1.2	ELA.RI.9-10.4	Identify descriptive words	Complete the meaning of a sentence	Determine the connotative meaning of words and phrases	Determine the figurative meaning of words and phrases	Determine how words and phrases impact text meaning
ELA.C1.2	ELA.RI.9-10.5	Draw conclusions from category knowledge	Answer who and what questions	Identify key details	Identify evidence for a claim	Differentiate between evidenced and non-evidenced claims
ELA.C1.2	ELA.RI.9-10.8	Understand difference of perspective	Identify an author's points	Identify the evidence for a claim	Analyze an argument and determine evidence for it	Determine how structure contributes to claims
ELA.C1.2	ELA.RL.9-10.1	Identify concrete details in a familiar story	Answer questions by referring to a text	Cite textual evidence for explicit information in text	Discriminate between explicit and implicit citations	Determine a narrative's explicit meaning
ELA.C1.2	ELA.RL.9-10.2	Identify the forward sequence in a familiar routine	Identify main idea	Identify details related to the theme of a story	Recount events contributing to the theme using details	Recount main events related to the theme
ELA.C1.2	ELA.RL.9-10.4	Identify descriptive words	Identify the words or phrases to complete a literal sentence	Determine the meaning of idioms and figures of speech	Determine the meaning of words and phrases	Determine the meaning and impact of words and phrases

Figure 1. Example diagnostic score report, summarizing the student's profile of mastery. Green shading indicates a skill mastered in the content standard (Essential Element). Gray shading in the Essential Element column indicates the student did not assess on that standard.



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 | approaching the target | at target | advanced

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 a category

The student identifies text elements by:

- identifying details in a familiar text

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

Individual Student Year-End Report
Performance Profile 2017-18

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.

Figure 2. The Performance Profile report includes the subject performance level, performance level descriptors, and conceptual area bar graphs summarizing the percent of skills mastered in each area of related standards.