Teacher Use of Score Reports for Instructional Decision-Making: Preliminary Findings

Amy K. Clark

Meagan Karvonen

Russell Swinburne Romine

Neal M. Kingston

ATLAS

University of Kansas

Author Note

Paper presented at the 2018 annual meeting of the National Council on Measurement in Education, New York, NY. Correspondence concerning this paper should be addressed to Amy Clark, ATLAS, University of Kansas, akclark@ku.edu.

Abstract

Consequential validity evidence is collected to evaluate the extent that assessment results are used as intended. Large-scale summative assessment results are typically used for program evaluation and resource allocation purposes; however, stakeholders increasingly desire results from large-scale K-12 assessments that inform instruction. Because large-scale summative results are typically delivered after the end of the school year, teacher use of results is reserved for the subsequent academic year. To evaluate use of summative score reports to inform instruction, a series of teacher interviews and focus groups were conducted with seventeen teachers in three states. Teachers were asked to describe how they used summative results from the 2016-2017 administration of a large-scale alternate assessment system in the subsequent academic year. Interview and focus group transcripts were coded for themes related to when and how score reports are delivered; how teachers use results to plan instruction, formulate IEP goals, and create instructional groupings; how teachers talk to parents about results; and what resources best support their use of score reports.

Keywords: score reporting, large-scale assessment, summative assessment, consequential evidence, instruction

Teacher Use of Summative Score Report for Instructional Decision Making: Preliminary Findings

Since the passage of the No Child Left Behind Act, states are required to administer educational assessments for accountability purposes. In addition to inclusion in accountability models, states and districts often use aggregated results for program evaluation and resource allocation purposes. While the results serve these purposes well, one of the historic limitations of large-scale assessments is that results have limited utility for informing instructional planning and decision making. With recent measurement and technological advancements resulting in assessments that are increasingly tailored to the student, including diagnostic assessments and computer adaptive measures, stakeholders desire results that are similarly nuanced and can be used to inform instruction.

When considering intended uses of large-scale assessment results, consideration must also be given to the reporting timeline. Because large-scale assessments are typically used for summative purposes, individual student score reports often are not delivered or made available to parents and teachers until after the academic year has concluded. In order for large-scale assessment results to be informative to instruction, they must then be actionable in the subsequent academic year, when students have already advanced a grade level and are being taught the academic content standards for the next grade. To support this use, score reports must not only communicate information about knowledge, skills, and abilities from the tested grade, but also be informative enough to facilitate teacher connection to current grade-level content.

As part of an assessment's validity framework, intended uses of assessment results should be defined a priori and described in program technical documentation. The *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014) state that evidence collected to evaluate an assessment's validity argument should include, among other sources of evidence, consequential evidence to evaluate the extent that assessment results are used for intended purposes. To the extent that assessment programs intend for results to be used to inform instruction or to discuss results with parents, evidence should be collected to evaluate the extent that teachers do so. To this end, evidence

is needed to evaluate how teachers use summative student score reports from the prior academic year to inform subsequent year instructional practice.

Context

Reporting

Guidelines for score reporting for educational assessments are presented in the *Standards* (AERA et al., 2014) and in the literature (e.g., Zenisky & Hambleton, 2012) to ensure appropriate interpretation and use of results. The *Standards* indicate "Score information should be communicated in a way that is accessible to persons receiving the score report. Empirical research involving score report users can help improve the clarity of reports (p. 200; AERA et al., 2014)." To support accurate interpretation and use, summative score reports should be designed with end users in mind, including teachers.

Goodman and Hambleton (2004) reviewed existing score reports and interpretive guides from approximately 15 programs to identify reporting strengths and weaknesses. They identified several strong reporting features, including summarizing assessment results in multiple forms, such as narrative and graphics; customizing reports for the intended audience; and providing interpretive guides. They also noted weaknesses, such as omission of the purpose of the assessment, how results should be used, precision of measurement, and definition of key terms; the use of statistical jargon; and dense, hard to read reports. These findings indicate the importance of balancing the inclusion of relevant information while also providing a report that is easy to read and understand to support stakeholders in interpreting results and using the information as intended.

Zenisky & Hambleton (2012) further summarized best practices for developing score reports, including a process for developing reports that involves stakeholder input, evaluation of prototypes, and refinement of reports over time. However, the authors re-emphasized the finding from Ryan (2003) regarding a need for additional research examining how practitioners actually use score reports once they are distributed. Extending their process, feedback on actual score report use should similarly be used to evaluate score report utility and make modifications to better support stakeholder interpretation and use.

Use of Summative Results

There is a limited body of research on how educators use results to inform instruction. Yeh (2006) conducted interviews with teachers and administrators on the use of results from federally mandated statewide assessments. Findings indicated that 56 of the 61 interviewees (92%) were concerned that reports provided inadequate diagnostic information about student knowledge, skills, and understandings. Interviews also noted that because results were from the prior year, they were less informative to the current year's instruction.

In a survey regarding teacher use of summative results, Hoover and Abrams (2013) asked teachers to indicate how they typically used data from various types of summative assessments, including teacher-made and district administered benchmark exams. Teachers reported most frequently evaluating aggregate results by examining the mean or mode, and less frequently disaggregating results for student subgroups or content standards. While the survey did not evaluate teacher use of state-mandated accountability measures, the results indicate that teachers did not use results in ways that would likely facilitate instructional practice, such as informing specific plans for instruction regarding student intervention or enrichment, or planning instructional groupings based on similar results across students.

Diagnostic Assessments

As stated, stakeholders increasingly desire assessment results that are useful to instruction. To facilitate thoughtful instructional planning and goal setting, providing fine-grained results beyond a performance level or raw or scale score is desired. Diagnostic assessments (e.g., Leighton & Gierl, 2007) have emerged as a measurement approach that can provide more detailed results to stakeholders due to scoring examinees on a series of attributes or skills measured by the assessment. Assessment items are associated with one or more skills, and a diagnostic scoring model (e.g., Bradshaw, 2017; Rupp, Templin & Henson, 2010) is used to provide the probability that a test taker mastered the attribute being measured. Because scoring is provided at the attribute level, diagnostic score reports provide more fine-grained information than is available for assessments providing a single raw or scale score value.

While it is widely acknowledged that diagnostic assessments hold more possibility for fine-grained reporting, there is limited evidence available evaluating diagnostic score report interpretation and use. Sinharay and Haberman (2009) urged that evidence must be collected to evaluate the extent that such additional reporting information actually results in improved practices related to instruction and meeting student needs. This caution reiterates the importance of collecting consequential evidence related to how score report results are actually used once distributed to parents and teachers.

Reporting for Alternate Assessments

Alternate assessments based on alternate achievement standards (AA-AAS) are designed for the approximately one percent of students with the most significant cognitive disabilities (SCD) who cannot access general education assessments even with accommodations. While AA-AAS allow students to be included in statewide accountability metrics, parents and teachers similarly desire useful information summarizing student performance, particularly when considering IEP goal development and instructional planning that is often one-on-one. Past limitations of AA-AAS score reports in particular have included limited information to guide changes in instructional practice (Nitsch, 2013) and evidence that teachers have not systematically used AA-AAS results when assessing progress or deciding what to teach after students have mastered academic skills (Karvonen, Wakeman, Moody, & Flowers, 2013).

AA-AAS score reports additionally provide unique opportunities for evaluation of use. While general education students often transition to a new teacher as they advance grades, teachers of students with significant cognitive disabilities often teach the same students year to year. These teachers' experiences provide a unique context for evaluating the use of summative results in the following academic year because teachers often have the benefit of already knowing the student's academic history firsthand, and possibly having administering the assessment the previous year.

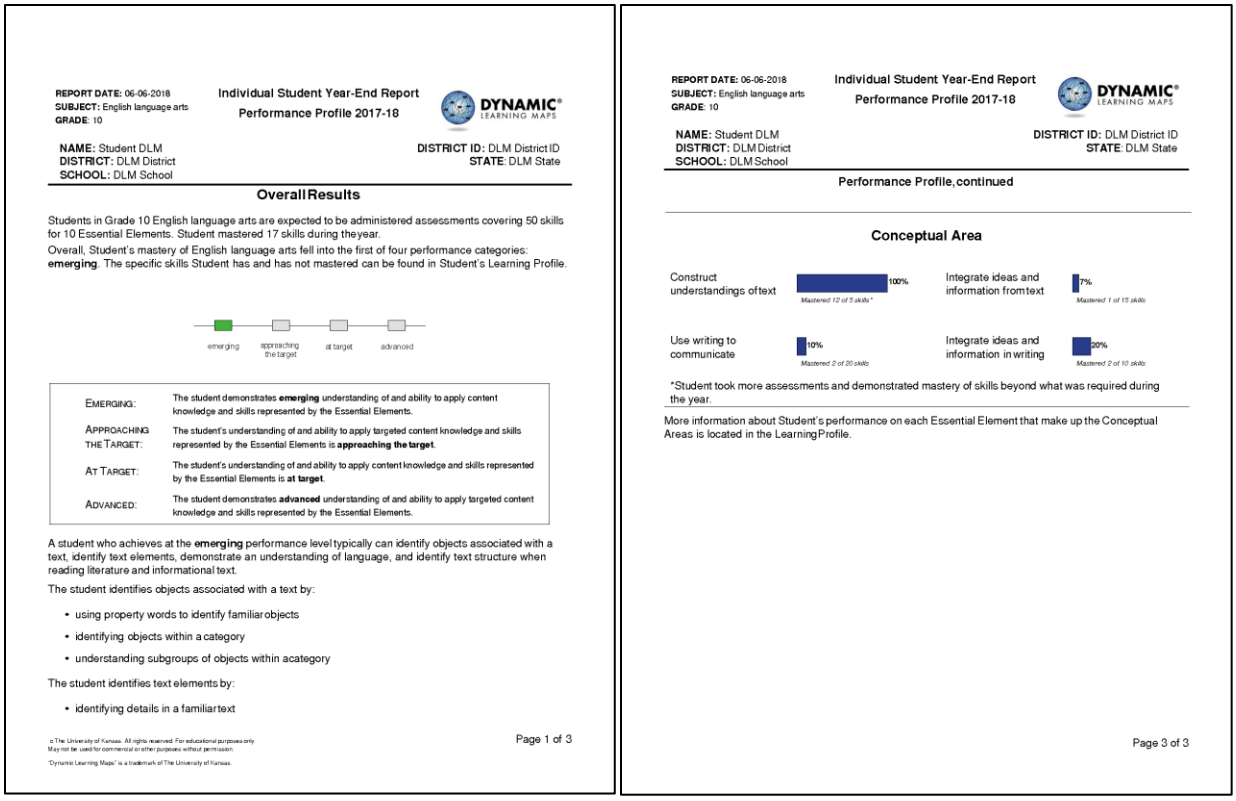
Dynamic Learning Maps Alternate Assessments

Dynamic Learning Maps (DLM) Alternate Assessments are available to students with the most significant cognitive disabilities in a seventeen-state consortium. DLM assessments are calibrated and scored using diagnostic classification modeling, and report performance based on mastery of many

discrete skills rather than a traditional raw or scaled score measuring a single latent trait. To meet the needs of the diverse student population, alternate content standards are available for assessment at five levels, called linkage levels, reflecting varied complexity from the grade-level target, including three precursor skills and one successor. Mastery of these skills serve as the basis of student assessment results.

States participating in the DLM Consortium can choose from either a through-course assessment model or a year-end assessment model. The through-course model includes instructionally embedded assessments on content standards of the teachers' choosing, from within blueprint constraints. Each spring, a short assessment of a subset of previously administered standards is used to update performance. Summative reporting is based on all responses collected throughout the year. The year-end assessment model, in contrast, bases summative reporting off only a summative spring assessment, which covers a fixed blueprint.

Because intended uses of DLM results include informing instructional planning, communicating student performance to parents, making programmatic decisions, and for inclusion in state accountability models, DLM reports were designed to provide actionable information to guide instructional decisions while also being appropriate for accountability purposes. Because more items are administered in the through-course assessment model than in the year-end assessment model, the through-course model score reports contain more information. Both assessment models receive a high-level Performance Profile summarizing results, displayed in Figure 1. The Performance Profile aggregates mastery information across the content standards into results reported for collections of related content, called conceptual areas, and the subject overall. The report summarizes results via the student's performance level, performance level descriptors typical of students achieving at that performance level, and conceptual area bar graphs summarizing the percent of skills the student mastered in each area. Students in states participating in the through-course assessment model additionally receive a fine-grained Learning Profile, displayed in Figure 2. The Learning Profile summarizes the mastery classifications for the five levels of skills available for each content standard, with text for each linkage level describing what is measured.



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 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

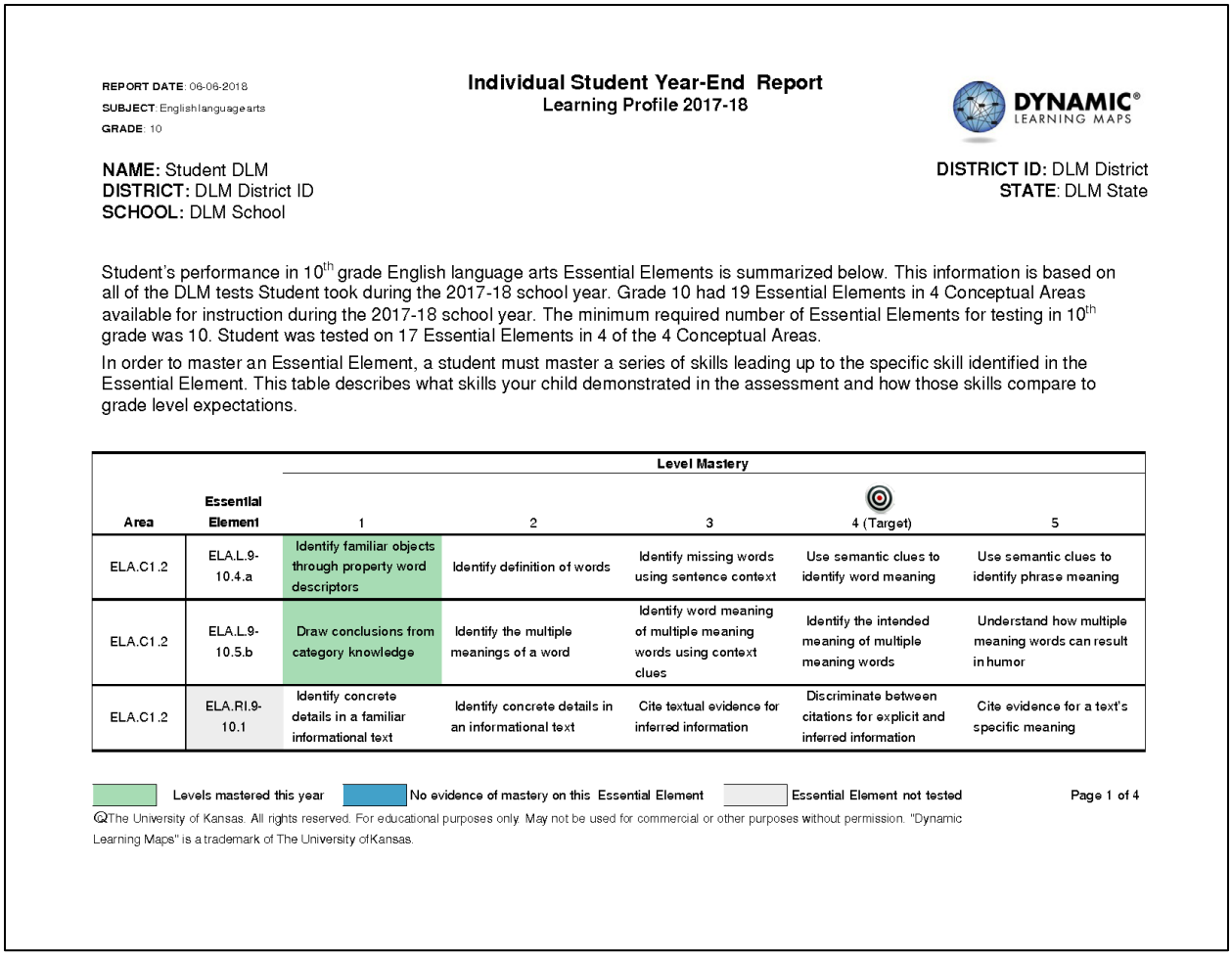
Performance Profile, continued

Conceptual Area

*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 1. Performance Profile report delivered to all states. Results include performance level, performance level descriptors, and conceptual area bar graphs summarizing the percent of skills mastered in each area.



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

Figure 2. Learning Profile report delivered to states participating in the through-course assessment model. Shading indicates skills mastered for the five linkage levels available for each "Essential Element" content standard.

Each report was first developed by DLM staff based on relevant research literature and refined through multiple rounds of focus groups conducted with educators and parents. Previous research has documented interpretability of the score report prototypes (Clark, Karvonen, Kingston, Anderson, & Wells-Moreaux, 2015), preliminary evidence of how teachers evaluate score report contents (Karvonen, Clark, & Kingston, 2016), and the impact of score report interpretation resources on teachers' understanding of report contents (Karvonen, Swinburne Romine, Clark, Brussow, & Kingston, 2017).

The current paper builds on this previous line of research by presenting results from teacher interviews and focus groups regarding how teachers actually use score reports to guide instructional decision-making. The study is designed to answer the following research questions:

1. How and when do teachers receive DLM summative score reports?
2. How do teachers use score reports to inform instructional decision-making?
3. How do teachers talk to parents about score reports?
4. Are there additional resources teachers need to support their use of score reports for instructional decision-making?

Methods

Teacher interviews and focus groups were conducted during spring 2018 with teachers recruited by consortium state partners from a sample of the 17 states in the DLM Consortium.

Participants

State partners recruited teachers to participate in small, virtual focus groups and interviews. State partners were provided with recruitment materials to distribute to potentially eligible teachers. Because the current study focused on use of reports in the subsequent academic year, several eligibility criteria were included. To participate, teachers must have indicated they:

1. currently teach one or more students who would take DLM assessments in 2017-2018,
2. received DLM 2017 summative score reports for their 2017-2018 students, and
3. used the DLM 2017 reports during the 2017-2018 academic year.

Interested teachers were asked to complete a Qualtrics survey listing their background information and responding to the three eligibility questions. A total of 135 teachers responded to the survey. Of those, 40 responded "yes" to all three eligibility questions and were contacted to set up a time to participate. Of those contacted, 17 participated in focus groups and interviews, including five teachers in through-course assessment model states, and 12 in year-end model states. Because of attrition challenges between scheduling and conducting phone calls, the number of participants per call ranged from one to five. This resulted in several focus groups being conducted as one-on-one interviews; they are

collectively referred to as focus groups for the remainder of the paper. All participants were compensated $50 for their time and contributions.

The 17 participating teachers represented 3 states and mostly self-reported as white (*n* = 13) and female (*n* = 13). Teachers taught in a range of settings, including rural (*n* = 2), suburban (*n* = 9), and urban (*n* = 5). Teachers reported a range of teaching experience by subject and for students with SCD, as indicated in Table 1, with most teaching more than one subject, and spanning all tested grades 3-12. Teachers indicated they taught between 1 (*n*=3) and 15 or more (*n* = 2) students currently taking DLM assessments, with most indicating they had between 2-5 students taking DLM assessments (*n* =8).

Table 1

Participating Teachers' Years of Experience per Subject and Population

Years	ELA	Math	Science	Students with SCD
1-5	4	4	5	6
6-10	4	5	3	4
11+	7	5	6	5

Data Collection

Participants were notified of focus group scheduling via email. The invitation included an informed consent document to sign and return and an example score report PDF. Because of the differences in score reports by assessment model, focus groups were conducted separately by model, with participants receiving an example report reflective of their own score reports (i.e., with or without the Learning Profile). The example score report was provided for one grade and subject. The provided example additionally had red overlay boxes with labels indicating the parts of the report (e.g., performance levels, descriptors, conceptual area graphs) to support orientation to the report, in the event someone needed to refer to a part of the report during the focus group discussion.

Focus groups were conducted virtually using Zoom conferencing software. Where possible, participants were encouraged to use video conferencing to facilitate participation. All focus groups were recorded for subsequent creation of verbatim transcripts. Each focus group began with a brief summary of

purpose, review of the informed consent document, brief review of score report contents, and an introduction of focus group participants. Following this introduction, focus groups followed a semi-structured format. Guiding questions were available to facilitate discussion and prompt for more information related to the four research questions. Participants were asked to describe receiving score reports, and how they use reports for their current students in planning or implementing instruction in any of the tested subjects (English language arts, mathematics, and/or science). Participants also described how they discuss reports with parents and shared information about available or desired resources to support their score report use.

Data analysis included the use of a coding protocol to determine how participants used DLM 2016-2017 summative score reports during the 2017-2018 academic year. Transcripts were coded according to descriptive themes related to the four research questions, including how reports were received; use for instruction, including planning, IEP goal development, and instructional groupings; talking with parents; and resources for parents, teachers, and districts. This paper summarizes preliminary findings, with a more intensive, multi-rater coding procedure and codebook to follow.

Results

Findings are summarized for each of the four research questions.

Receiving Reports

The DLM Consortium makes individual student score reports available at the state level and to district test coordinators in an online platform. States and districts have differing policies regarding distribution of reports to schools, teachers, and parents at the local level. Despite responding in the affirmative to the eligibility questions around score report use, several teachers indicated in the focus group that the score reports they received were actually different than those discussed in the focus group. One teacher from a through-course model state indicated only receiving the Learning Profile portion of the score report.

All teachers who received reports indicated receiving them in the fall, typically from their district or building test coordinator. Several mentioned their district test coordinator delivered reports at an

annual meeting that also included required annual test administrator training. Fewer indicated receiving the reports as part of a meeting intended to discuss results. Others reported receiving only an email to notify them score reports were ready, with no additional explanation or interpretive materials provided. A review of consortium practice indicated 11 states made reports available to building test coordinators, while only three states made individual student score reports available in the online portal to teachers. With increased budgetary and resource demands at the local level, including concerns over printing costs, additional information is needed to understand why more states and districts do not make reports available to teachers in the online platform.

Using Reports to Inform Instruction

Participant discussion revealed varying levels of utility of student score reports for planning instruction, with particular differences observed by grade level. Teachers of elementary and middle school students whose accountability requirements included annual assessment found reports to be more useful 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 years prior, particularly for their eleventh grade students who only had 8th grade reports available. Teachers also noted that often the curriculum in 12th grade, as students prepare to transition, was markedly different from the 11th grade curriculum, and therefore results from the prior year were not as useful. In contrast, elementary and middle school teachers, and especially those who instruct the same students year to year, reported much more utility in using reports for planning instruction, specifying IEP goals, and planning instructional groupings.

Instructional Planning. Teachers in the through-course assessment model described their processes for using fine-grained results summarized in the Learning Profile 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 varied based on student needs. For some students, teachers described focusing less on skills that had already been mastered to provide greater breadth of instruction and assessment; for others, they described prioritizing

the next level of skill acquisition within a similar standard to provide greater depth of instruction and assessment.

Teachers in the year-end assessment model shared more varied feedback for using reports to inform instructional planning. One strategy included using the performance level descriptors to know the skills typical of students in the performance level. Another strategy involved the use of the conceptual area bar graphs to know the percent of skills mastered in each area of related content standards. Finally, teachers described relating the information on the DLM score reports to information obtained from local assessment results as an additional source of information about student performance.

IEP Goals. Teachers in both groups described using reports to inform IEP goal planning, however, teachers who received the more fine-grained Learning Profile reported greater utility. As one teacher stated, "Their IEP goals are very similar to their linkage level [statement]. I can say, 'Hey, let's look at this linkage level and let's look at this target skill 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."

By contrast, teachers in year-end states receiving only the Performance Profile reported using the performance level descriptors as a general framework. One mentioned using the conceptual area bar graphs combined with results from a district assessment to frame IEP goals. The teacher stated, "We have a district assessment in the fall, winter, and spring, so in the fall, they provide a report and summary. I try to see if there is still a deficiency based on the DLM [results from] the spring and in the new report in the fall to see if that is an area that there's still a weakness, and if there is then that's definitely something I would spend more time on. That's more of how I create my goals using the concept." However, other teachers indicated the report was not specific enough to frame IEP goals, particularly because the year-end assessment model does not capture growth over time. These teachers reported using data from other progress monitoring and district tools to inform IEP goal development.

Instructional Groupings. As mentioned, participating teachers had a range of students taking DLM assessments, with some only having a single student. Teachers also indicated that while they may have more than one student, they might be in different grades or perform at very different levels based on unique cognitive and communication considerations. However, in instances where multiple students were assessed in the same grade, teachers talked about the benefit of being able to plan instructional groupings from reports. Teachers who received the Learning Profile mentioned using performance on the linkage levels to plan instruction for students working on the same skills across standards. Another teacher expressed a desire for an aggregated report that made instructional groupings more clear, particularly around standards and levels students were working on in common.

Talking with Parents

Teachers highlighted the importance of understanding the assessment and student results when talking to parents. As one teacher stated, "That first year… I wasn't able to give the parents a lot other than, 'Here's your score report,'" and indicating the performance level. In contrast, by the second year the teacher mentioned knowing more about the process of selecting academic content standards and levels as part of the through-course assessment model. She stated, "I know more about where they are going and what they're doing so I can share that with parents… This is the academic focus, this is what we're hoping they get out of reading that aligns with their IEP goals, which aligns with the DLM testing. It is a better conversation about why this testing format is."

For parents of students new to the DLM assessment system, teachers reported some confusion about the reports. "Parents seemed a little confused because they had never seen a report before. So I don't think they really knew exactly what they were looking at since it was something so new presented to them." The teacher went on to share, "We just went over exactly what was on the report step by step. I pointed out some of the IEP objectives and how they were related to what was on the report."

Most teachers reported that while their district shared a copy of the report to give to parents, they were not provided with the DLM Parent Interpretive Guide to accompany the report, and teachers were not aware it existed. Teachers in the year-end model, without the availability of the Learning Profile

summarizing the specific skills mastered, reported that parents seemed unsure how the student performance level was determined. As one stated, "The mathematical formula was not very cut and dry, so it was very difficult to explain it to them." While the Performance Profile contains narrative text in addition to the performance level graphic (shown in Figure 1), these teacher comments indicate the report likely needs to go further in explaining how the performance level was determined to be informative to parents.

Overall, teachers reported that, with a few exceptions, parents did not ask questions about the DLM assessment or score reports, so the extent of information parents received about the assessment was dependent upon what the teacher offered. As one teacher indicated, "Unfortunately, I just don't think that our parents know what to ask. They're not educated about the test. They only have the information that I give them and so, this year I was able to give them more, but will I be able to give them even more information at the end of the year when we transition their child off to middle school? Oh yeah, because I've looked at it better so I could give more information."

Resources

Findings related to resources include feedback on existing resources as well as resources teachers mentioned would be beneficial if made available. Feedback is summarized as resources for parents, teachers, and district education agencies.

Parents. Teachers noted that often parent-teacher conferences and IEP meetings inundate parents with information about their student from a variety of sources, including classroom observation data, IEP goals, summative testing results, and may include information from multiple staff members. Because these meetings may leave parents feeling overwhelmed, teachers suggested making resources available that could be introduced at the meeting but available on the website for parents after they have digested as well. They suggested a brief overview, such as a short video explaining the testing system and how student results are calculated. The site could also include resources for explaining results to parents, such as the Parent Interpretive Guide, and cheat sheets for how parents can tie academic content measured

by the assessment into their day-to-day interactions with their children, for example when they visit a grocery store.

Teachers. Participants described a number of resources that would be useful in supporting teachers in interpreting and using reports, including training opportunities and additional reporting. One suggestion was for a set of three teacher meetings each fall. The first would be a meeting for teachers to complete required test administrator training and receive their certification. At the second meeting, teachers would receive their summative score reports from the year prior and discuss how to read the score report and what the different sections contain. At the third meeting, teachers would be able to start planning instruction from the score report, including opportunities for cross-teacher collaboration, for example, a student's fifth grade teacher working with the student's sixth grade teacher to discuss the student's learning trajectory and begin planning instruction.

Teachers also expressed a desire for more information in aggregate form. As mentioned, teachers who worked in self-contained settings with larger caseloads expressed the desire for reports geared toward identifying instructional groupings, at the class level. These reports would provide them an easy way to quickly see students working on similar areas and aid in planning their instruction.

Districts. Many teachers highlighted the strong value of district-provided professional development activities. While many teachers do not currently receive professional development related to DLM assessments, several emphasized the strong benefit such opportunities would provide. Teachers suggested district trainings for interpreting results and for planning instruction, and also suggested providing district agencies with summary reports they could use to support teachers. They suggested these reports could be used to evaluate from a programmatic level if certain standards or conceptual areas were perhaps being covered less or that teachers may struggle to teach. By identifying these areas collectively, the participants suggested districts may be better equipped to point teachers toward already available resources or host trainings to address potential areas of challenge.

Discussion

Results from the DLM assessment system are intended for several uses, including inclusion in state accountability models, reporting results to districts, teachers, and parents, and for use in instructional planning and decision making. Evaluating actual use of assessments results is an integral part of an assessment's validity argument. Further, evaluation of teacher use of summative results, particularly in the subsequent academic year, addresses a gap in the literature regarding how teachers use score reports for instructional decision making. The study demonstrates the impact large-scale assessment results can have on improving teaching and learning for students and teachers beyond their inclusion in statewide accountability metrics and for evaluation at state and district levels.

Preliminary findings highlighted here also demonstrate the promise of DCM-based system score reports that provide fine-grained results. Feedback from teachers in the through-course model who received the Learning Profile reported more utility for results to inform instructional planning, IEP goal specification, and plans for instructional groupings. They also expressed fewer misconceptions and misunderstandings related to the scoring process and how performance levels in the Performance Profile were determined. Overall these findings suggest the potential that more detailed reports obtained from diagnostic platforms may support for informing instructional practice and communicating results to parents.

The results of the present study also highlight several unique challenges to alternate assessment systems. While the DLM Consortium pools sample sizes across participating states, within individual states sample sizes may still be low. Several teachers reported being the only teacher in their school or area administering DLM assessments. Additionally, because so few students take alternate assessments, state and district resources may be harder to obtain. In instances where schools or districts have limited time and resources, a single training per year may be all they can support and that training may prioritize assessment administration due to the consortium-wide requirement for annual test administrator training. While maintaining state and local control are important aspects of educational assessment, making resources available on a broader scale is one way to leverage consortium participation.

Beyond alternate assessment systems, districts face myriad resource challenges for supporting the various local and state-mandated assessments delivered in their schools. Particularly for measures that are mandated rather than locally selected, districts may experience challenges knowing what resources would be useful to teachers or how to explain the reporting for those measures. While teachers indicate a desire for more training, time and availability of staff to provide numerous training opportunities may be limited.

To address these challenges, large-scale assessment systems can make resources available to support the administration and interpretation of results with fidelity. This might include readily available materials districts can point schools towards 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 assessment results to have an important and meaningful impact on students and teachers in the classroom.

References

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington DC: American Educational Research Association.

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.

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.

Hoover, N. R., & Abrams, L. M. (2013). Teachers' instructional use of summative student assessment data. *Applied Measurement in Education, 26,* 219-231. doi: 10.1080/08957347.2013.793187

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., 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.

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

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

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.

Ryan, J. M. (2003). *An analysis of item mapping and test reporting strategies*. Greensboro, NC: SERVE.

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

Yeh, S. S. (2006). Reforming federal testing policy to support teaching and learning. *Educational Policy, 20*, 495-524. doi: 10.1177/0895904805284119

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.