

Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Copyright is retained by the first or sole author, who grants right of first publication to *Practical Assessment, Research & Evaluation*. Permission is granted to distribute this article for nonprofit, educational purposes if it is copied in its entirety and the journal is credited. PARE has the right to authorize third party reproduction of this article in print, electronic and database forms.

Volume 28 Number 6, May 2023

ISSN 1531-7714

Teacher Use of Score Reports for Instructional Decision-Making in the Subsequent Academic Year

Amy Clark, *University of Kansas: ATLAS*
Jennifer L. Kobrin, *University of Kansas: ATLAS*
Meagan Karvonen, *University of Kansas: ATLAS*
Ashley Hirt (*University of Kansas: ATLAS*)

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 usually 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, we conducted a series of teacher interviews and focus groups with 17 teachers in three states. Teachers were asked to describe how they used summative results from the previous administration of a large-scale alternate assessment system in the subsequent academic year. Interview and focus-group transcripts were coded and identified themes related to when and how score reports are delivered; how teachers use results to plan instruction, formulate goals for individualized education programs (IEPs), and create instructional groupings; how teachers talk to parents about results; and what resources best support their use of score reports. Findings demonstrate preliminary support for diagnostic score report use.

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

Introduction

Since the passage of the No Child Left Behind Act in 2001, states are required to administer educational assessments for accountability purposes. States and districts often use aggregated performance results for program evaluation and resource allocation. Although summative assessment results serve these intended uses well, one historic drawback of large-scale assessments is their limited utility for informing instructional planning and decision-making (e.g., Marion, 2018). Large-scale assessment score reports are often delivered after the academic year has concluded; to inform instruction and parent–teacher communication, they must be actionable in the subsequent academic year, when students have

advanced a grade and are being taught the subsequent grade’s content standards.

As part of an assessment’s validity framework, intended uses of results should be defined a priori and described in technical documentation. The *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA] et al., 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 results are used as intended. To the extent that assessment programs intend for results to inform instruction or foster parent–teacher communication, programs should evaluate the extent that reports are used in this way. To this end, evidence

is needed to evaluate how teachers use summative student score reports from the prior academic year to inform instructional practice and foster parent communication in the subsequent year and which score-reporting features teachers find most useful for these purposes.

Context

Score Reporting

Much of the score reporting (also known as results reporting; O'Donnell & Zenisky, 2020) literature focuses on design and development best practices to promote appropriate interpretation and use by various stakeholders (e.g., Slater et al., 2018; Zapata-Rivera, 2019; Zapata-Rivera & Katz, 2014). Seminal publications (e.g., Goodman & Hambleton, 2004; Hambleton & Zenisky, 2013; Zenisky & Hambleton, 2012) and guidelines in the field (e.g., AERA et al., 2014) describe procedures for developing score reports that involve steps for collecting stakeholder input, evaluating prototypes, and refining reports over time. They describe strong reporting features, including summarizing assessment results in multiple formats (e.g., narrative and graphics), customizing reports for the intended audience(s), and providing interpretive guides. They also describe practices to avoid, such as dense reports; use of statistical jargon; and omission of assessment purpose, intended uses, information regarding measurement precision, and definitions of key terms. Several researchers have described considerations for score-report design specific to parent audiences (e.g., Kannan et al., 2018, 2021; Rios & Ihlenfeldt, 2021) and how reporting can foster parent-teacher communication (Barber et al., 1992; Dries, 2014; Hopster-den Otter et al., 2017).

Although literature on score-report design and development is prevalent, it frequently mentions the need for additional research examining how educators and parents actually use reports after distribution (e.g., Kannan et al., 2018; O'Leary, 2017; Ryan, 2003; Slater et al., 2018; Zenisky & Hambleton, 2012). A recent review highlighted this discrepancy, noting that while studies of score-report properties and design are well represented, few studies examine practical interpretation and use of score reports, and more research is needed in this area (Gotch & Roberts, 2018). In instances where interpretation and use is

evaluated, findings sometimes indicate discrepancies between actual and intended uses (van der Kleij & Eggen, 2013). Collecting feedback on actual score-report use should be used to evaluate score-report utility and modify reports to better support stakeholder interpretation and use (Zenisky & Hambleton, 2012); the feedback can also contribute to validity evidence (MacIver et al., 2014; O'Leary et al., 2017; Tannenbaum, 2019).

Use of Assessment Results

Large-scale assessment programs typically describe several intended uses of summative assessment results. Examples include measuring annual achievement and progress; communicating performance information to various audiences; evaluating teacher, school, and district performance; and supporting instructional planning (e.g., Florida Department of Education, 2021; North Dakota Department of Public Instruction, 2022; NWEA, 2021; Smarter Balanced Assessment Consortium, 2020; West Virginia Department of Education, 2020). However, their validity evidence tends to skew toward evidence for accountability uses of large-scale assessment results with little on using results for informing instruction or to foster communication with parents. Some programs do not include informing instruction as an intended use and instead rely on formative and interim measures for instructional decision-making, which likely contributes to teacher perceptions that summative assessments are not useful to their practice (Olson & Jerald, 2020).

Programs in which results are intended to be instructionally useful are constrained by reports' being delivered too late in the year to inform the tested grade's instruction (Marion, 2018). Further, summative reporting tends to prioritize reporting overall achievement, which is useful for accountability reporting but less informative for instruction. Interviews on the use of results from federally mandated assessments revealed that many teachers and administrators (92%) were concerned that reports provided inadequate diagnostic information (Yeh, 2006), and teachers criticized the lack of meaningful information contained in score reports or the information's application to instruction (Bonner et al., 2018). Some assessment programs advocate for reporting subscores to provide finer grained, instructionally useful information to stakeholders, but test designs tend to limit the number of items

measuring each subscore, which introduces psychometric challenges for reporting results (e.g., low reliability; Sinharay et al., 2019) and subsequent interpretation.

Hoover and Abrams (2013) asked 656 teachers in a large suburban district to indicate how they typically used summative assessment data, including teacher-made and district-administered benchmark exams (but not state-mandated accountability measures). Teachers reported most frequently evaluating aggregate results by examining the mean or mode, and less frequently disaggregating results for student subgroups or content standards. They indicated that instructional changes resulting from summative data were limited and primarily geared toward future instruction, with minimal or no use of results to inform reteaching or teacher reflection on past instruction.

When reported effectively, summative assessment results can inform instructional practice in the subsequent academic year. For example, case study interviews with 101 teachers, principals, and district staff from six districts and one charter school found that teachers used student-achievement data to adjust instructional practices (Dembosky et al., 2005). Teacher responses indicated three levels of instructional adjustment: whole-class, small group, and individualized. Many teachers reported using results from the previous year's state standardized assessment to modify their instruction for the current year, identify content that challenged students, and adjust their whole-class instruction to focus on the areas in which students underperformed. The survey findings indicated that teachers most often used assessment data to organize differentiated small groups, particularly at the elementary level.

This nearly 20-year-old study, along with the 10-year-old Hoover and Abrams (2013) study, were the only ones we identified that examined use of summative results in the subsequent academic year. Related literature in data-based decision-making (also referred to as *data-driven decision-making*) shows that teachers tend to have low confidence in their ability to use data to inform instructional decisions (Bettesworth et al., 2008; Dunn et al., 2013a, 2013b; Gummer & Mandinach, 2015); and similarly notes the dearth of research documenting how teachers use data to inform instruction (e.g., Farley-Ripple et al., 2021; Ruhter & Karvonen, in press).

Interpretation and use challenges are exacerbated when reporting achievement for students with significant cognitive disabilities who take alternate assessments based on alternate achievement standards (AA-AAS). Parents of these students have historically not received score reports, or the reports did not contain useful information (Nitsch, 2013). Reports tended to show high rates of proficiency despite students' cognitive disabilities and need for alternate achievement standards (Thurlow et al., 2012), or reflected unstable performance over time (Thurlow et al., 2012; Wu et al., 2021); both challenges further reduced the usability of the results. Score reports also had limited information to guide changes in instructional practice (Nitsch, 2013), and teachers did not systematically use AA-AAS results when assessing progress or deciding what to teach after students had mastered academic skills (Karvonen et al., 2013). These collective findings across multiple literature bases emphasize the need for consequential evidence about how test data are used to inform instructional practice, including talking with parents about assessment results.

Diagnostic Assessments

Providing fine-grained results beyond a performance level, raw score, or scale score is desirable to facilitate thoughtful instructional planning and goal setting. Recent measurement and technological advancements support providing assessments that are increasingly tailored to the student and can provide teachers with useful information to inform learning. Diagnostic assessments (e.g., Leighton & Gierl, 2007) have emerged as a measurement approach that can provide more-detailed results by scoring examinees on a series of attributes or skills. Assessment items are associated with one or more skills, and a diagnostic classification model (DCM; e.g., Bradshaw, 2017; Rupp et al., 2010) is used to determine the probability that a test taker mastered the attributes being measured. Because scoring is provided at the attribute level, diagnostic score reports provide finer grained information than is available for assessments providing a raw- or scale-score value. These assessments can also meet state accountability needs when mastery results are aggregated to report overall achievement (Clark et al., 2017).

A widely acknowledged strength of diagnostic assessments is the ability to provide fine-grained reporting (e.g., Rupp et al., 2010). The literature

provides several example reports and interpretive guides for diagnostic assessment systems (Clark et al., 2015; Jurich & Bradshaw, 2014; Roberts & Gierl, 2010; Rupp et al., 2010) and demonstrates how assessment literacy can influence interpretation of diagnostic assessment results (Clark et al., 2022). Several conference papers have described teacher feedback on score reports from assessments scored with diagnostic classification modeling (Feldberg & Bradshaw, 2019; Karvonen et al., 2016, 2017), but there are limited instances of the operational use of DCMs to provide student results (Sessoms & Henson, 2018). We know of no published papers evaluating teachers' actual use of DCM-based score reports from an operational assessment program.

Research Questions

The current paper addresses the gap in research evaluating teacher use of summative score reports, particularly to inform instruction in the subsequent academic year and to foster parent communication, in the context of a diagnostic assessment system that provides fine-grained results in addition to reporting overall achievement in the subject. Because score-report distribution and associated interpretation resources can affect teachers' ability to use results as intended, we also evaluated how teachers receive reports. We addressed three specific research questions.

1. How, and with what resources, do teachers receive summative score reports?
2. How do teachers use score reports to inform instructional decision-making in the subsequent academic year?
3. How do teachers talk to parents about score reports?

Methods

Because stakeholder perspectives are critical to our understanding of score-report use, we designed the study to collect feedback from teachers via interviews and focus groups (Trainor & Graue, 2014).

Study Context

Dynamic Learning Maps (DLM) alternate assessments are administered in 21 states to students with significant cognitive disabilities. DLM

assessments are calibrated and scored using a DCM, and they report performance as mastery of many discrete skills rather than a traditional raw or scale score measuring a single latent trait (DLM Consortium, 2018). To meet the needs of the diverse student population, alternate content standards are available for assessment at five levels, reflecting varied complexity from the grade-level target, including three precursor skills and one successor. Mastery of these skills is the basis of student assessment results.

States administering DLM assessments can choose from either a through-course or year-end assessment model. The through-course model evaluates student mastery using instructionally embedded assessments of content standards that teachers choose from within blueprint constraints. Summative reporting is based on all responses collected throughout the year. In contrast, the year-end assessment model determines summative mastery from a spring assessment covering a fixed blueprint.

DLM results are intended for use in state accountability models; other intended uses include informing instruction, communicating student achievement to parents, and making program decisions. To support these uses, DLM score reports were designed to provide actionable information to guide instructional decisions while also reporting overall performance in each subject (English language arts, mathematics, and science). At the time of this study, the through-course assessment model administered more items per content standard than the year-end model, resulting in more-detailed information in the through-course-model score reports. Both assessment models generate a high-level Performance Profile summarizing overall achievement (Figure 1). The Performance Profile aggregates mastery information across the content standards, reporting mastery for collections of related content strands, called *conceptual areas*, and for the subject overall. The report summarizes results with a performance level, performance level descriptors summarizing skills typical of students achieving at that performance level, and conceptual-area bar graphs summarizing the percentage of skills a student mastered in each area. Students in states adopting the through-course assessment model also received a fine-grained Learning Profile (Figure 2). The Learning Profile summarizes the mastery classifications for the five levels of skills available for each content standard, with text

describing each level. This difference in reporting format across models provides a unique opportunity to understand how educators used the differing contents.

Each score report was first developed by DLM staff using relevant research literature and refined through input from state education agency staff and multiple rounds of educator and parent focus groups. Previous research has documented interpretability of the score-report prototypes (Clark et al., 2015), preliminary evidence of how teachers evaluate score-report contents (Karvonen et al., 2016), and the impact of score-report interpretation resources on teachers' understanding of report contents (Karvonen et al., 2017).

The DLM Consortium delivers individual student score reports through an online platform. Reports are

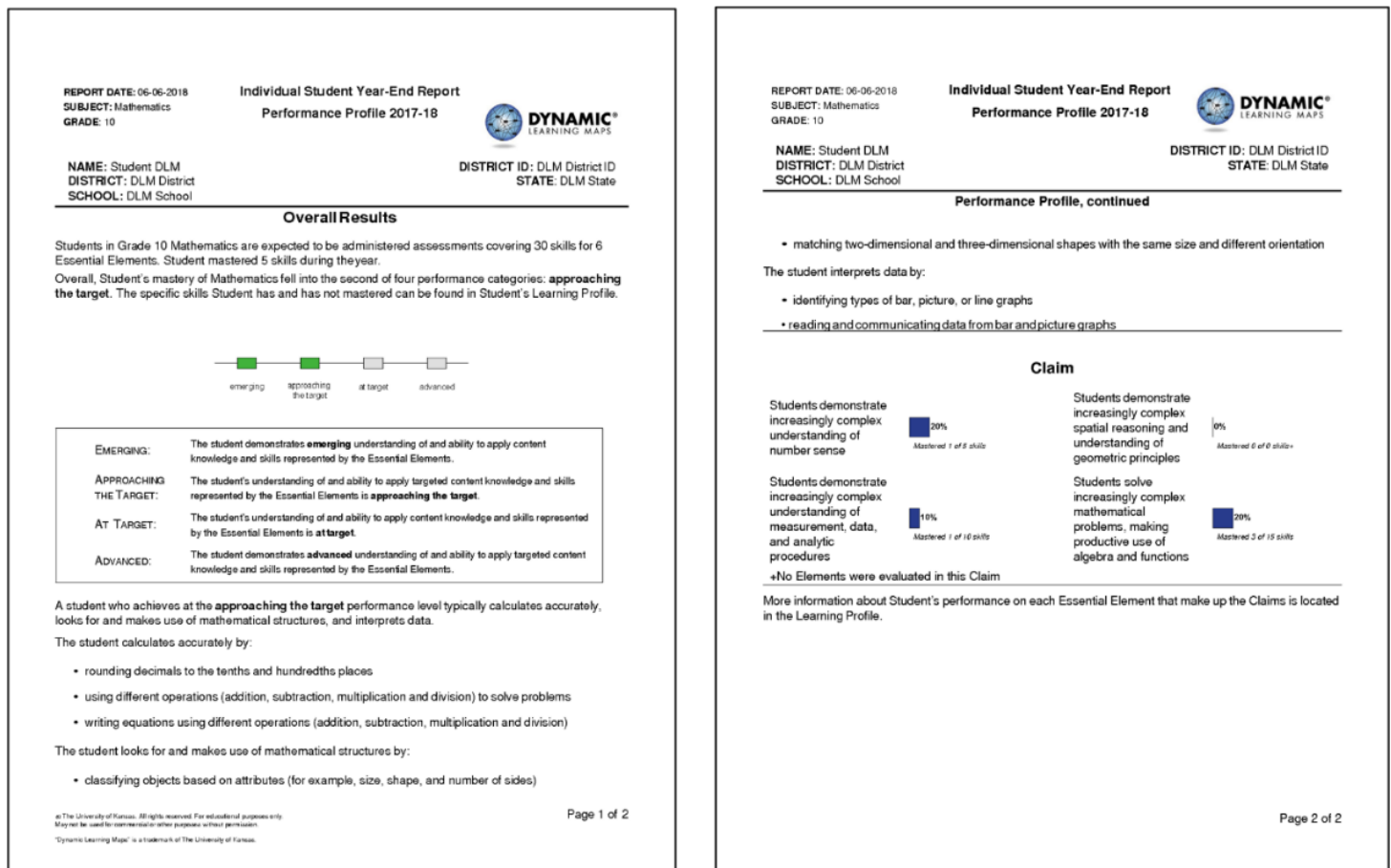
delivered to state agency staff and to district test coordinators. States and districts have differing policies regarding local distribution of reports to schools, teachers, and parents. Some make reports available in the online system (e.g., to district or building coordinators only, versus teachers), whereas others only mail printed copies.

Participants

State education agency staff from both models distributed recruitment materials to teachers. Because the study focused on report use in the subsequent academic year, teachers must have met all of the following eligibility criteria to participate:

1. taught one or more students who took DLM assessments
2. received prior year summative score reports for their current students

Figure 1. Example Performance Profile Report Delivered to All States




Note. Results include performance level, performance level descriptors, and conceptual-area bar graphs summarizing the percentage of skills mastered in each area.

Figure 2. Example Learning Profile Report Delivered to States Participating in the Through-Course Assessment Model

REPORT DATE: 06-06-2018
SUBJECT: Mathematics
GRADE: 10

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

Individual Student Year-End Report
Learning Profile 2017-18



DISTRICT ID: DLM District ID
STATE: DLM State

Student's performance in 10th grade Mathematics 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 26 Essential Elements in 4 Claims available for instruction during the 2017-18 school year. The minimum required number of Essential Elements for testing in 10th grade was 6. Student was tested on 6 Essential Elements in 3 of the 4 Claims.

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.

Claim	Essential Element	Level Mastery				
		1	2	3	4 (Target)	5
M.C1	M.N-CN.2.a	Recognize set, subset, and separateness	Combine; Use repeated addition and multiplication	Multiply by 1-5 and 10; add within 20	Apply properties of addition and multiplication to solve problems.	Explain properties of multiplication/addition
M.C1	M.N-CN.2.b	Recognize separateness and objects in a set	Recognize unit; know place value, 1 ten = 10 ones	Add/subtract decimals with digits in the tenths place	Solve word problems with rational numbers	Solve multi-step word problems
M.C1	M.N-CN.2.c	Recognize separate objects	Recognize unit; know place value, 1 ten = 10 ones	Multiply twodecimals with digits in the tenth place	Solve word problems with rational numbers	Solve multi-step word problems

Levels mastered this year
 No evidence of mastery on this Essential Element
 Essential Element not tested

©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.
Page 1 of 5

Note. Shading indicates skills mastered for the five levels available for each Essential Element content standard.

- used the prior year's summative score reports during the current academic year

Interested teachers were asked to complete a Qualtrics survey describing their background information and responding to the three eligibility questions. In total, 170 teachers responded to the survey. Of those, 40 responded affirmatively to all three eligibility questions and were contacted to set up a time to participate. Of those contacted, 17 teachers 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 the focus groups, the number of participants per interview ranged from one to five, resulting in several focus groups becoming one-on-one interviews; they are collectively referred to as focus groups throughout.

The 17 participating teachers represented three states and self-reported as White ($n = 13$), Hispanic/Latino ($n = 1$), Native American ($n = 1$), and Black ($n = 1$). Participants included female ($n = 14$) and male ($n = 3$) teachers. Teachers taught in rural ($n = 2$), suburban ($n = 9$), and urban ($n = 5$) settings. Teachers reported a range of teaching experience by subject and for students with significant cognitive disabilities

(Table 1), with most teaching more than one subject, spanning all tested grades 3–12. Teachers indicated they taught between one ($n = 3$) and 15 or more ($n = 2$) students taking DLM assessments, and most indicated they had between two and five students taking DLM assessments ($n = 8$).

Data Collection

Participants were notified of focus-group scheduling via email. The invitation included an informed consent document to sign and return and a PDF example of a score report. Because of the differences in score reports by assessment model, focus groups were conducted separately by model, with participants receiving an example report that reflected the reports they would have received for their students (i.e., with or without the Learning Profile). The example score report was provided for one grade and subject. The example featured red overlay boxes with labels that indicated the parts of the report (e.g., performance levels, descriptors, conceptual-area graphs), 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. When possible, participants were encouraged to enable video to facilitate participation. Focus groups were recorded for subsequent creation of verbatim transcripts. Each focus group began with a summary of purpose, review of informed consent and score-report contents, and introduction of participants. Focus groups then followed a semistructured interview format. Participants were asked to describe how they received score reports and how they use reports for their current students. Participants also described how they discuss reports with parents and shared information about available or desired resources to support their score-

report use. Participants were compensated \$50 USD for their time and contributions.

Data Analysis

Recordings were professionally transcribed and de-identified for coding. We developed an initial codebook using a combination of deductive and inductive approaches (e.g., Linneberg & Korsgaard, 2019; Thomas, 2006). We identified general coding categories based on descriptive themes related to the research questions, including how reports were received and available resources to support use; use for instruction, including planning, development of goals for individualized education programs (IEPs), and instructional groupings; and talking with parents. We then conducted a microanalysis, (i.e., a close reading and detailed line-by-line analysis of the transcripts; Corbin & Strauss, 2014) to identify specific coding categories. This resulted in four general coding categories and 21 specific coding categories. Teacher responses to a question or follow-up probe were each considered a separate segment for coding, and segments could be assigned one or more codes.

Four researchers coded the transcripts independently and then met to discuss discrepancies and reach consensus. The codebook was revised and updated as needed to refine, combine, or add codes. Transcripts coded early were recoded with the final codebook. To analyze the data, we conducted a content analysis of coded segments to identify patterns and themes.

A variety of methods were used to establish rigor (e.g., credibility, trustworthiness, transferability; Amankwaa, 2016; Eldh et al., 2020; Lincoln & Guba, 1985). We included diverse participant perspectives as a form of triangulation (Patton, 1999). We asked clarifying questions during focus groups and

Table 1. Participating Teachers' Years of Experience by Subject and Population

Years	English language arts	Mathematics	Science	Students with significant cognitive disabilities
1–5	4	4	5	6
6–10	4	5	3	4
11+	7	5	6	5

paraphrased statements to confirm researchers' understanding. We also relied on collaborative work, involving multiple researchers in interviewing and coding processes, and we used a codebook to promote consistency in coding. We include direct participant quotes in our results for transparency in the conclusions we draw.

Positionality

The research team included the four coauthors and a support staff member. Three of the coauthors hold doctorates in educational psychology and research, and one holds a doctorate in musicology but works in educational research. The support staff member assisted with coding transcripts and pulling code reports from Dedoose qualitative software. All members of the research team worked on the DLM project. We acknowledge our program affiliation warrants interest in its success, but our research agenda also includes opportunities to collect feedback to inform continuous improvements.

Results

Findings are summarized for the three research questions.

Receiving Reports and Supplementary Resources

Participants described a variety of ways they received reports, reflecting variability in state and district policies and procedures regarding report distribution. Some participants received reports in the online portal, while others received them via email or mail or as a hard copy during an in-person meeting. Despite affirmatively responding to the eligibility questions (i.e., they had received DLM score reports and used them), two teachers said that the score reports they received were different from those discussed in the focus group, indicating they had not, in fact, received the summative score reports. Additionally, one teacher from a through-course model state indicated receiving only the Learning Profile portion of the report rather than both parts.

Teachers typically received reports in the fall from their district or building test coordinator. One teacher from a through-course model state shared, "I'm real on

top on that one, and I'm calling to make sure I get them as quick as possible before the beginning of the year. Before we start testing for sure."¹ Several participants said their district test coordinator delivered reports at an annual meeting that also included completing required annual test-administrator training. Fewer teachers 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, or they received the mailed set of reports, with no additional explanation or interpretive materials provided.

The amount of support participants received as part of score-report delivery varied. One through-course model teacher described annual meetings during which teachers used the prior year's summative score reports to select content standards for instruction and assessment in the current year. Another teacher described sitting with the district test coordinator to review the results together. Another teacher shared,

Our testing coordinator is awesome. She helps provide us with examples of what we should be teaching based on the score reports, and what we could improve on and what we should introduce or work on more, so we're provided with a lot of assistance from our testing coordinator.

Other teachers received less support. Some described local training that focused more on the design and delivery of the assessment and less on how to interpret and use the results they receive. As one teacher stated,

What I need to know is how to apply that to my students specifically going forward to next year, and if I don't have the same students, I still can apply it, you know, to the next student who's similar.

Some teachers had to seek out resources individually. One stated, "Just on my own, I feel I've found more information than what's actually been given to me by a coordinator." Others perceived that the lack of support was because alternate assessments generally receive less attention at building, district, and state levels. As one teacher stated, "There's not a ton of guidance necessarily. Part of it could be because I'm

¹ The fall through-course assessment window runs from September to December.

the only one in my building, kind of off in an oasis.” Another compared the level of support received to what is provided for general education assessments:

[General assessment] is a movie star, and DLM is an assistant associate producer in [our state]. There’s nothing out there for DLM. Like zero—they don’t care about it, it’s an afterthought. So what happens is, the more information that’s provided on the report, the more information parents and teachers have.

These comments underscore the variability in the level of support teachers receive at the district and building level, which in turn can affect teachers’ familiarity with report contents and how comfortable they are interpreting and using results.

Some teachers suggested supports they would like to receive to facilitate interpretation and use. One described wanting three meetings each fall. During the first meeting, teachers complete required test-administrator training. At the second meeting, teachers receive summative score reports and discuss how to read the report, including the contents of each section. At the third meeting, teachers start planning instruction from the report, including opportunities for cross-teacher collaboration (e.g., 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). Other teachers expressed a desire for delivery of results in aggregate form. For instance, teachers who worked in self-contained settings with larger caseloads wanted reports geared toward identifying instructional groupings. These reports would provide an easy way to identify students working on similar areas and aid in planning their instruction.

Many teachers highlighted the strong value of district-provided professional development for supporting report interpretation and use. While many teachers did not currently receive professional development on DLM assessment results, several emphasized the benefit such opportunities would provide. Teachers suggested district training for interpreting results and for planning instruction and providing district staff with summary reports they could use to support teachers. They suggested these reports could be used to evaluate, from a programmatic level, certain standards or conceptual areas to emphasize or that teachers may struggle to

teach. By identifying these areas collectively, the participants suggested that districts may be better equipped to direct teachers to available resources or offer trainings to address areas of challenge.

Although the DLM program offers various resources to support score interpretation and use, participants’ use of those materials varied. Materials include stakeholder-specific interpretive guides for district users, teachers, and parents, along with short videos on the DLM website. Some teachers had accessed these resources, but others were not aware of them and others had difficulty finding the materials. Given that many teachers said they had to find resources on their own, rather than receiving them from building or district staff, this feedback underscores the need for materials that are easy to locate and navigate.

Using Reports to Inform Instruction

Participant discussion revealed varying levels of utility of student score reports for planning instruction in the subsequent year. In elementary and middle grades, where assessments are required annually, teachers found reports to be more useful than in high school, where students are typically assessed in one 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 11th-grade students whose most recent report was from eighth grade. Teachers also noted that the curriculum in 12th grade, as students prepare for post-school transition, was often markedly different from the 11th-grade curriculum, so results from the prior year were not as useful. In contrast, elementary and middle school teachers, especially those who instruct the same students across years, reported much more utility in using reports for planning instruction, specifying IEP goals, and evaluating results across students.

Instructional Planning. Teachers in the through-course assessment model described their processes for using the fine-grained results 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 in the current grade’s content standards.

We were taught in our training, these are our guides.... It tells me exactly what they’ve mastered... what we need to focus on based on the blueprint

and then based on what their scores were from last year. So I look at the blueprint for the new year and then I look at their scores from last year, and I say, “What in the blueprint says that we need to test on this year and what in their overall results?” I look at what they scored last year in each subject and what the gaps are, and that’s the areas we’re going to focus on this year academically.

Having informative score reports can help when planning instruction, especially for students with significant cognitive disabilities, because districts do not always provide curriculum for these students. In those instances, the structure of the Learning Profile (i.e., one row per content standard, with columns for the five levels) can help teachers determine gaps and next steps for learning. One teacher said,

This is what we’re going to teach; it’s our curriculum guide. They don’t make a curriculum for [these students] so what we do is..., I look at their scores. Let me give you an example. Their Essential Element is MD.3. Recognize some objects and separate objects. The student, that’s as far as they got, they earned a 1 [first level], they didn’t reach target [on this Essential Element]. So if [I] want to get them on target for that task, then I need to work on recognizing enclosures [second level], explaining unit squares area to area [third level]. So I use that; this is my guide—this is where we need to get him. If that’s [a] task that they’re tested on this year. I want to get them on target. What am I going to have to do to get there?

Teachers’ prioritization of instructional content varied. Some teachers described prioritizing more-complex content associated with a similar standard to provide greater depth of instruction and assessment. Others described focusing less on content standards that had already been mastered to provide greater breadth. One teacher stated, “If I have a choice between the [standard] he’s already targeted and the one that he has not, then I’ll pick the one he has not.”

Teachers in the year-end assessment model (i.e., without the Learning Profile) shared more-varied feedback for using reports to inform instructional planning. One strategy included using the performance level descriptors to identify skills typically mastered by students achieving at the performance level. One teacher said,

I think the most valuable [part] of this score report is actually the grade- and content-specific performance level descriptors... because at least it tells you a student who achieves the approaching [level] typically can do this, this, this, and that. That helps a lot.

Another teacher shared, “I looked at the areas where our students couldn’t do [the skill]. I looked at those [descriptors,] and I picked out a few to focus on.” In the absence of the finer-grained Learning Profile, educators relied on the performance level descriptor statements of typical mastery to inform instructional planning.

Another strategy involved the use of the conceptual-area bar graphs to determine the percentage of skills mastered for related content standards. As one teacher stated, “If there’s a deficit, based on the conceptual area, that might be an area I want to hone in on and create my goals for.” Another teacher described using the conceptual-area information in a similar way:

For my one student, you can see [in which conceptual areas] she really didn’t score that high, and that’s what we really worked on this year, hoping that the test coming up soon will show more growth... and see if we can make it go further.

Because the conceptual areas are the same across grades and provide continuity across the grade-specific standards, teachers used the percentages to identify areas of strength and need for planning subsequent instruction.

Finally, teachers described relating the information on the DLM score reports to information obtained from class and district assessment results as an additional source of information about student performance. One teacher described use of other assessment data:

We have a district assessment in the fall, winter, and spring, so in the fall, they provide a report and summary, almost at what level the kids could be instructed at. So 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.

In this example, the teacher used both sets of results to identify instructional priorities in the subsequent year rather than relying on a single data source.

IEP Goals. Teachers in both groups described using reports to inform IEP goal planning, but teachers who received the fine-grained Learning Profile reported greater utility. One participant said she used the level statements on the Learning Profile (i.e., five levels for each standard): “I really feel like this holds kids to a higher standard. I think it keeps teachers from writing cop-out goals.” Another teacher shared that she used the leveled skills in the Learning Profile to specify her goal, making the grade-level target the expectation and the sequential levels the short-term objectives.

By contrast, without the Learning Profile, teachers in year-end states described using the report for IEP development differently. One shared using the reports as a starting point for IEPs:

It helps us to plan her educational plan or educational binder for the year, and it helps me also to do her progress reports and goals in her IEP because then I have something more concrete, I feel, to go off of.

Another described more specifically using the conceptual-area-mastery information “kind of as a guideline for when I create my goals in my IEPs.” Yet another teacher described combining the results with other sources of information to inform IEP development.

I wouldn’t say I necessarily go off of this, and say, “Oh you know, based on the DLM, the student can do A, B, and C; therefore, we’re putting it in.” I mean, a lot of times it is something I do put in, just because, knowing the students, the DLM does give a pretty accurate depiction of where the kid is at, so, I mean, there is some connection, but I wouldn’t just solely use the DLM results and put that into the present level of performance.

As with instructional planning, teachers reported using data from other progress monitoring and district tools to inform IEP goal development and evaluation.

Evaluating Across Students. As stated earlier, the number of students for whom participants administered DLM assessments varied, with some administering to just one student. Teachers also reported that, while they may have multiple students,

they may be in different grades or perform at different levels based on unique cognitive and communication considerations. However, in instances where multiple students were assessed in the same grade, teachers described the benefit of being able to evaluate learning across students and use score reports to plan instructional groupings. Teachers who received the Learning Profile mentioned using mastery on the five 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 clearer, particularly identifying common standards and levels across students. A high school teacher described using collective student results to improve her instruction more generally.

I looked at it across all the students.... I think we use the score reports to improve ourselves a lot. I think they’re very helpful for us as teachers as well, to kind of look back and say, “Okay, what can we do better? What can we do different?”

Talking to Parents

Teachers’ communication with parents about summative score reports also varied. In some districts, teachers were responsible for providing the reports to parents, while other districts centrally distributed reports to parents.

Teachers highlighted the importance of understanding the assessment and student results when talking to parents. As one teacher stated, during her first year of administration, “I was able to read the score report to [parents], but I couldn’t tell them much other than the right-there information.” In contrast, by the second administration year, the teacher knew more about the process of selecting academic content standards and levels as part of the through-course assessment model:

I know more about where [students] 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 [this way].

Teachers described parents of students new to the DLM assessment having confusion about the reports. One said, “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 continued, “We just went over exactly what was on the report, step by step.” Another teacher echoed this sentiment.

It was really their first experience getting that information, too, so it was kind of new to all of us.... So I think they were okay with it because they didn't know any better, I didn't know any better—we were kind of learning together.

Most teachers reported that, while their district provided the report to give to parents, it did not provide the DLM Parent Interpretive Guide to accompany the report, and many teachers were not aware it existed. Lacking the Learning Profile that summarizes the specific skills mastered, teachers in the year-end model reported that parents seemed unsure how the performance level was determined. As one teacher stated, “The mathematical formula was not very cut and dry, so it was very difficult to explain it to [parents].” While the Performance Profile contains narrative text in addition to the performance level graphic (shown in Figure 1), these teacher comments indicate that, to be informative to parents, the report likely needs to further explain how the performance level was determined.

Overall, teachers reported that, with a few exceptions, parents did not ask questions about the 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 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.

Although many teachers described using parent-teacher conferences or IEP meetings to deliver reports to parents, they also said these meetings inundate parents with materials. Parents receive a range of information about their student, including classroom observation data, IEP goals, and summative testing results, and may receive 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 also available on a website for later access. They suggested a brief overview, such as a short video, describing the assessment and how results are calculated. The site could also include additional resources, such as the Parent Interpretive Guide, and quick guides for how parents can connect academic content measured by the assessment to day-to-day interactions with their children (e.g., visiting the grocery store).

Discussion

This study expands the limited literature (Dembosky et al., 2005; Hoover & Abrams, 2013) on use of score reports in a subsequent academic year. While this study demonstrates how large-scale assessment results can inform teaching and learning beyond inclusion in accountability metrics, it also underscores the importance of report delivery to teachers and parents, the contents of the reports (as emphasized by Zenisky & Hambleton, 2012 and others), and resources that support their use.

For teachers to use score reports as intended, they need to know not only how to interpret their students' score reports, but also how to transform data from the reports into information to make instructional decisions (Gummer & Mandinach, 2015). They also need to be able to make connections to content knowledge and pedagogical content knowledge to inform instructional planning (Blanc et al., 2010; Schildkamp et al., 2019). This and other studies (e.g., Farley-Ripple & Buttram, 2014; Gesel et al., 2021; Ruhter & Karvonen, in press) revealed that teachers have varying levels of support (professional development or other resources) to develop this knowledge.

We observed some variability in teachers' sophistication in using score reports. While this may be attributable to differences in their data-based decision-making, content knowledge, or level of support, it could also be attributed to differences across the two assessment models (i.e., varying score-report specificity). The limited sample size of this study does not support conclusions about model-based differences, but additional research could continue to explore the differences in complexity of score-report

use from the level of information presented. For instance, through-course model teachers who received the Learning Profile tended to describe more utility and specificity for results informing instructional planning, IEP goal specification, and instructional groupings. These teachers also expressed fewer misconceptions and misunderstandings related to the scoring process and how performance levels in the Performance Profile were determined. Interestingly, neither group described using the overall performance-level text or graphic as a useful source of information, despite its prevalence as a reporting mechanism for large-scale assessments.

This study also provides support for diagnostic score reporting. Teacher descriptions of using the Learning Profile and aspects of the Performance Profile derived from mastery-based scoring (e.g., percentage of skills mastered by conceptual area) are unique strengths of mastery-based reporting. Although more research is needed, these findings underscore the potential that fine-grained diagnostic reports have for informing instructional practice and communicating results to parents.

This study also identified important challenges districts and teachers encounter with score-reporting resources and training, some of which may be unique to AA-AAS and some may be more common. Districts must support the various local and state-mandated assessments delivered in their schools; staff may experience challenges in knowing what resources would be useful to teachers or how to explain the reporting, particularly for assessments that are mandated rather than locally selected. These challenges may be exacerbated for AA-AAS, where fewer students take the assessments, teachers may be isolated, and district staff may be less prepared to support them.

While teachers indicated a desire for more training, the time and availability of building and district staff to provide further opportunities may be limited. When schools or districts have limited time and resources, a single training per year may be all they can feasibly support, and that training may prioritize assessment administration. Supporting district staff in knowing the critical information about the assessment and score interpretation may aid them in expanding the scope of training materials. This may include creating materials for use during professional development and

professional learning community activities. Districts could leverage these resources during district-provided in-service training to better equip teachers to use results to inform instruction and share pertinent information with parents. Districts also could support teachers in using information from multiple data sources, as some focus-group participants described. Given the lack of interim assessments specifically designed for students with significant cognitive disabilities (Boyer & Landl, 2021; Browder et al., 2021; Lazarus et al., 2021), additional assessment development is likely needed to support special educators in equitably combining multiple sources of evidence, as their peers do.

Beyond expanding district support, promoting awareness of available resources is important so teachers can access the information they need. Making resources easily accessible on a broader scale may be one way to support teachers and parents in using score reports. However, we also recognize that it can be challenging for teachers to find time to find and use additional materials. Therefore, it is imperative for test developers to create reports that communicate the critical information needed to interpret and use results.

Implications for Research and Practice

We share four implications for research and practice identified from this study.

1. Reports should be readily available. For stakeholders to interpret and use results, timely distribution is necessary. This study encountered recruitment challenges in identifying teachers who had actually received their summative score reports. Of 170 teachers who indicated interest in this study, only 40 said they had received and used score reports; of those interviewed, two indicated the reports they received had actually not been the summative reports we provided in our materials. While this may be a product of historical AA-AAS challenges (e.g., Nitsch, 2013), it may also extend to general education contexts and may contribute to the perception that large-scale assessments are not useful to instructional practice (Olson & Jerald, 2020).

One possible way to combat these challenges is online report distribution and score-report dashboards (e.g., Zapata-Rivera & Katz, 2014). In addition to making report distribution

timelier for educators, online report distribution may also address budgetary and resource demands at the local level, including concerns about printing costs. A review of DLM Consortium practice at the time of the focus groups indicated that 11 states made PDF reports available online to building test coordinators, while only three states made PDF reports available online to teachers. As of 2022, the number of states allowing teachers access to online score reports increased to 14. As programs consider shifting toward online distribution and reporting dashboards, they should consider best practices, including making the materials easy to find and broadly communicating their availability. Additional research on online and interactive reporting will further inform this work.

2. Reports should be interpretable. The score-report literature emphasizes how report contents directly contribute to validity evidence for the assessment program (Hambleton & Zenisky, 2013; Tannenbaum, 2019). Feedback from the range of stakeholders that receive score reports is critical to making reports interpretable. In this study, feedback on the language included in the report prompted an external review to eliminate overly complex or confusing language in an effort to better support teachers' and parents' interpretation and use of the report. Although literature is available on supporting score-report interpretation, the examples are largely based on score reports derived from classical and item-response-theory scoring models. As diagnostic assessments continue to gain prevalence in the field, additional research on report interpretability is needed, particularly because of the differences in how diagnostic results are produced and reported.
3. Programs should support stakeholders in using results. As identified in this study, stakeholders may receive the score report in isolation, without supplementary interpretive guides to aid their use. To the extent possible, the score report itself should include enough information for recipients to meaningfully use the results. During the focus groups, teachers

emphasized their desire for more resources and their challenges related to finding and accessing supplemental materials to support interpretation. As a result of this study, DLM developers added text and a link to the footer in the score reports directing readers to a website with additional resources that support interpretation and use. Teachers also shared a desire for more district-level training and support. Their feedback underscores that simply designing high-quality reports is not enough. Programs should also invest in collecting feedback to identify the types of supports users want; future research could explore design and development of these supports as well.

4. Continually evaluate score-reporting goals for areas of improvement. This study demonstrates the value of collecting feedback on reports over time. While valuable information can be collected during score-report design phases, programs should not conclude that the reports are final once early feedback is collected. Programs should commit to continual improvement of reports, including evaluating whether needs have changed over time. For instance, during design phases for DLM assessments, year-end-model states originally prioritized short assessments with a limited number of items per standard, resulting in the distribution of only the Performance Profile portion of score reports. However, upon observing the benefit of the Learning Profile, these states decided to expand the number of items measuring each standard to be able to reliably report skill mastery and provide the Learning Profile. This change, along with the others mentioned earlier, underscore the importance of continual and ongoing evaluation of score reports. We strongly urge other assessment programs to collect stakeholder feedback continually—through surveys, focus groups, or other methods—to understand how results are used, to identify potential challenges and their solutions, and to disseminate research findings to the field to continue advancing the quality of score reports we make available.

Limitations

We acknowledge several limitations in the present study. As stated earlier, recruitment proved challenging and resulted in a small sample size. Further, because the inclusion criteria required teachers to have received and used reports, the participants may have differed from their peers who did not receive and use reports. We acknowledge that study participants may have been more invested in score-report use and may have described more sophisticated use of reports than other teachers in the population. However, given the sparse literature evaluating teachers' score reports use (Gotch & Roberts, 2018) and the lack of literature evaluating diagnostic score report use, we believe the findings presented here make an important contribution to the literature.

References

- Amankwa, L. (2016). Creating protocols for trustworthiness. *Qualitative Research Journal of Cultural Diversity*, 23(3), 121–127.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Barber, B. L., Paris, S. G., Evans, M., & Gadsden, V. L. (1992). Policies for reporting test results to parents. *Educational Measurement: Issues and Practice*, 11(1), 15–20. <https://doi.org/10.1111/j.1745-3992.1992.tb00222.x>
- Bettesworth, L. R., Alonzo, J., & Duesbery, L. (2008). Swimming in the depths: Educators' ongoing effective use of data to guide decision making. In T. J. Kowalski & T. J. Lasley (Eds.), *Handbook on data-based decision making in education* (pp. 286–303). Routledge.
- Blanc, S., Christman, J. B., Liu, R., Mitchell, C., Travers, E., & Bulkley, K. E. (2010). Learning to learn from data: Benchmarks and instructional communities. *Peabody Journal of Education*, 85(2), 205–225. <https://doi.org/10.1080/01619561003685379>
- Bonner, S. M., Torres Rivera, C., & Chen, P. P. (2018). Standards and assessment: Coherence from the teacher's perspective. *Educational Assessment, Evaluation and Accountability*, 30(1), 71–92. <https://doi.org/10.1007/s11092-017-9272-2>
- Boyer, M., & Landl, E. (2021). *Interim assessment practices for students with disabilities* (Research Brief NCEO Brief No. 22). National Center on Educational Outcomes; National Center for the Improvement of Educational Assessment. <https://nceo.umn.edu/docs/OnlinePubs/NCEOBrief22.pdf>
- 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. Wiley.
- Browder, D., Lazarus, S., & Thurlow, M. (2021). *Alternate interim assessments for students with the most significant cognitive disabilities* [NCEO Brief No. 23]. National Center on Educational Outcomes. <https://nceo.umn.edu/docs/OnlinePubs/NCEOBrief23.pdf>
- Clark, A. K., Karvonen, M., Kingston, N., Anderson, G., & Wells-Moreaux, S. (2015, April 15–19). *Designing alternate assessment score reports that maximize instructional impact*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL, United States.
- Clark, A. K., Nash, B., & Karvonen, M. (2022). Teacher assessment literacy: Implications for diagnostic assessment systems. *Applied Measurement in Education*, 35(1), 17–32. <https://doi.org/10.1080/08957347.2022.2034823>
- 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. <https://doi.org/doi:10.1111/emip.12162>
- Corbin, J., & Strauss, A. (2014). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Sage.
- Dembosky, J. W., Pane, J. F., Barney, H., & Christina, R. (2005). *Data driven decision making in Southwestern Pennsylvania school districts* [Working Paper WR-326-HE/GF]. Rand Education. https://www.rand.org/pubs/working_papers/WR326.html

- Dries, S. D. (2014). *The influence of parent portal access on student efficacy and parental involvement* (Publication No. 2076) [Doctoral dissertation, Seton Hall University]. Seton Hall University Dissertations and Theses (ETDs). <https://scholarship.shu.edu/dissertations/2076/>
- Dunn, K. E., Airola, D. T., Lo, W. J., & Garrison, M. (2013a). Becoming data driven: The influence of teachers' sense of efficacy on concerns related to data-driven decision making. *Journal of Experimental Education*, 81, 222–241.
- Dunn, K. E., Airola, D. T., Lo, W. J., & Garrison, M. (2013b). What teachers think about what they can do with data: Development and validation of data driven decision making efficacy and anxiety inventory. *Contemporary Educational Psychology*, 38(1), 87–98. <https://doi.org/10.1016/j.cedpsych.2012.11.002>
- Dynamic Learning Maps Consortium. (2018). 2017–2018 *Technical manual update—Year-end model*. University of Kansas: Accessible Teaching, Learning, and Assessment Systems (ATLAS). https://dynamiclearningmaps.org/sites/default/files/documents/publication/2017-2018_YE_Technical_Manual_Update.pdf
- Eldh, A. C., Årestedt, L., & Berterö, C. (2020). Quotations in qualitative studies: Reflections on constituents, custom, and purpose. *International Journal of Qualitative Methods*, 19, 1609406920969268. <https://doi.org/10.1177/1609406920969268>
- Farley-Ripple, E. N., & Buttram, J. L. (2014). Developing collaborative data use through professional learning communities: Early lessons from Delaware. *Studies in Educational Evaluation*, 42, 41–53. <https://doi.org/10.1016/j.stueduc.2013.09.006>
- Farley-Ripple, E. N., Jennings, A., & Jennings, A. B. (2021). Tools of the trade: A look at educators' use of assessment systems. *School Effectiveness and School Improvement*, 32 (1), 96–117. <https://doi.org/10.1080/09243453.2020.1777171>
- Feldberg, Z., & Bradshaw, L. (2019, April 4–8). *Reporting results from diagnostic classification models for teachers*. Paper presented at the annual meeting of the National Council on Measurement in Education, Toronto, Canada.
- Florida Department of Education. (2021). *Florida statewide assessments 2020–2021 technical report: Vol. 5*. https://www.fldoe.org/core/fileparse.php/5663/urlt/V5FSA_TechRep2021.pdf
- Gesel, S. A., LeJeune, L. M., Chow, J. C., Sinclair, A. C., & Lemons, C. J. (2021). A meta-analysis of the impact of professional development on teachers' knowledge, skill, and self-efficacy in data-based decision-making. *Journal of Learning Disabilities*, 54(4), 269–283. <https://doi.org/10.1177/0022219420970196>
- 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. https://doi.org/10.1207/s15324818ame1702_3
- Gotch, C. M., & Roberts, M. R. (2018). A review of recent research on individual-level score reports. *Educational Measurement: Issues and Practice*, 37(3), 46–54. <https://doi.org/10.1111/emip.12198>
- Gummer, E. S., & Mandinach, E. B. (2015). Building a conceptual framework for data literacy. *Teachers College Record*, 117, 1–22. <https://doi.org/10.1177/016146811511700>
- 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. I. C. Hansen, N. R. Kuncel, S. P. Reise, & M. C. Rodriguez (Eds.), *APA handbook of testing and assessment in psychology, Vol. 3: Testing and assessment in school psychology and education* (pp. 479–494). American Psychological Association. <https://doi.org/10.1037/14049-023>
- Hoover, N. R., & Abrams, L. M. (2013). Teachers' instructional use of summative student assessment data. *Applied Measurement in Education*, 26, 219–231. <https://doi.org/10.1080/08957347.2013.793187>
- Hopster-den Otter, D., Wools, S., Eggen, T. J., & Veldkamp, B. P. (2017). Formative use of test

- results: A user's perspective. *Studies in educational evaluation*, 52, 12–23. <https://core.ac.uk/download/pdf/82492526.pdf>
- Jurich, D. P., & Bradshaw, L. P. (2014). An illustration of diagnostic classification modeling in student learning outcomes assessment. *International Journal of Testing*, 14(1), 49–72. <https://doi.org/10.1080/15305058.2013.835728>
- Kannan, P., Zapata-Rivera, D., & Bryant, A. D. (2021). Evaluating parent comprehension of measurement error information presented in score reports. *Practical Assessment, Research, and Evaluation*, 26(12). <https://doi.org/10.7275/RGWG-T355>
- Kannan, P., Zapata-Rivera, D., & Leibowitz, E. A. (2018). Interpretation of score reports by diverse subgroups of parents. *Educational Assessment*, 23(3), 173–194. <https://doi.org/10.1080/10627197.2018.1477584>
- Karvonen, M., Clark, A. K., & Kingston, N. (2016, April 7–11). *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, United States. https://dynamiclearningmaps.org/sites/default/files/documents/presentations/NCME_2016_DLM_Score_Report.pdf
- Karvonen, M., Swinburne Romine, R., Clark, A. K., Brussow, J., & Kingston, N. (2017, April 26–30). *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, United States. https://dynamiclearningmaps.org/sites/default/files/documents/presentations/NCME_2017_Score_Reports.pdf
- 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. <https://doi.org/10.1080/09362835.2012.747184>
- Lazarus, S. S., Hinkle, A. R., Liu, K. K., Thurlow, M. L., & Ressa, V. A. (2021). *Using interim assessments to appropriately measure what students with disabilities know and can do: Advisory panel takeaways and NCEO recommendations* (NCEO Report #427) (ED13064). ERIC; National Center on Educational Outcomes. <https://files.eric.ed.gov/fulltext/ED613064.pdf>
- Leighton, J. P., & Gierl, M. J. (2007). *Cognitive diagnostic assessment for education: Theory and applications*. Cambridge University Press.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage.
- Linneberg, M. S., & Korsgaard, S. (2019). Coding qualitative data: A synthesis guiding the novice. *Qualitative Research Journal*, 19(3), 259–270. <https://doi.org/10.1108/QRJ-12-2018-0012>
- MacIver, R., Anderson, N., Costa, A.-C., & Evers, A. (2014). Validity beyond interpretation: A user validity perspective beyond the test score. *International Journal of Selection and Assessment*, 22(2), 149–164. <https://doi.org/10.1111/ijsa.12065>
- Marion, S. F. (2018). The opportunities and challenges of a systems approach to assessment. *Educational Measurement: Issues and Practice*, 37(1), 45–48. <https://doi.org/10.1111/emip.12193>
- Nitsch, C. (2013). *Dynamic Learning Maps: The Arc parent focus groups*. The Arc. <https://dynamiclearningmaps.org/sites/default/files/documents/publication/TheArcParentFocusGroups.pdf>
- No Child Left Behind Act of 2001, P.L. 107-110, 20 U.S.C. § 6301 et seq. (2002).
- North Dakota Department of Public Instruction. (2022). *North Dakota state assessment for science 2020–2021, Vol. 1: Annual technical report*. https://ndsa.portal.cambiumast.com/-/media/project/client-portals/north-dakota/pdf/2022/v1_north-dakota-state-assessment-for-science-technical-report_2020-2021.pdf
- NWEA. (2021). *Spring 2021 NSCAS Phase 1 Pilot ELA, mathematics, and science technical report*. <https://files.eric.ed.gov/fulltext/ED617116.pdf>
- O'Donnell, F., & Zenisky, A. L. (2020). Digital module 21: Results reporting for large-scale assessments.

- Educational Measurement: Issues and Practice*, 39(4), 137–138. <https://doi.org/10.1111/emip.12408>
- 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. <https://doi.org/10.1111/emip.12141>
- Olson, L., & Jerald, C. (2020). *The big test*. FutureEd. https://www.future-ed.org/wp-content/uploads/2020/04/TheBigTest_Final.pdf
- Patton, M. Q. (1999). Enhancing the quality and credibility of qualitative analysis. *Health Services Research*, 34(5 part 2), 1189–1189.
- Rios, J. A., & Ihlenfeldt, S. D. (2021). State assessment score reporting practices for English learner parents. *Educational Measurement: Issues and Practice*, 40(3), 31–41. <https://doi.org/10.1111/emip.12424>
- 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>
- Ruhter, L., & Karvonen, M. (in press). The impact of professional development on teachers’ use of data-based decision making for students with significant cognitive disabilities. *Remedial and Special Education*.
- Rupp, A., Templin, J., & Henson, R. (2010). *Diagnostic measurement: Theory, methods, and applications*. The Guilford Press.
- Ryan, J. M. (2003). *An analysis of item mapping and test reporting strategies*. SERVE.
- Schildkamp, K., Smit, M., & Blossing, U. (2019). Professional development in the use of data: From data to knowledge in data teams. *Scandinavian Journal of Educational Research*, 63(3), 393–411.
- Sessoms, J., & Henson, R. A. (2018). Applications of diagnostic classification models: A literature review and critical commentary. *Measurement: Interdisciplinary Research and Perspectives*, 16(1), 1–17. <https://doi.org/10.1080/15366367.2018.1435104>
- Sinharay, S., Puhari, G., Haberman, S. J., & Hambleton, R. K. (2019). Subscores: When to communicate them, what are their alternatives, and some recommendations. In D. Zapata-Rivera (Ed.), *Score reporting research and applications* (pp. 35–49). Routledge.
- Slater, S., Livingston, S. A., & Silver, M. (2018). Score reports for large-scale testing programs: Managing the design process. In D. Zapata-Rivera (Ed.), *Score reporting research and applications* (1st ed., pp. 91–106). Routledge.
- Smarter Balanced Assessment Consortium. (2020). *2018-19 summative technical report*. <https://technicalreports.smarterbalanced.org/2018-19-summative-report/book/>
- Tannenbaum, R. J. (2019). Validity aspects of score reporting. In D. Zapata-Rivera (Ed.), *Score reporting research and applications* (pp. 9–18). Routledge.
- Thomas, D. R. (2006). A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation*, 27(2), 237–246. <https://doi.org/10.1177/1098214005283748>
- Thurlow, M., Altman, J., & Vang, M. (2012). *2008-2009 APR snapshot #3: AA-AAS participation and performance*. University of Minnesota, National Center on Educational Outcomes. <https://nceo.info/Resources/publications/APR-snapshot/AAAAS/default.html>
- Trainor, A. A., & Graue, E. (2014). Evaluating rigor in qualitative methodology and research dissemination. *Remedial and Special Education*, 35, 267-274. <https://doi.org/10.1177/0741932514528100>
- van der Kleij, F. M., & Eggen, T. J. H. M. (2013). Interpretation of the score reports from the Computer Program LOVS by teachers, internal support teachers and principals. *Studies in Educational Evaluation*, 39(3), 144–152. <https://doi.org/10.1016/j.stueduc.2013.04.002>
- West Virginia Department of Education. (2020). *West Virginia general summative assessment 2018-2019: Vol. 1, annual technical report*. https://www.wvde.us/wp-content/uploads/2020/06/WVGSA_Volume-1_SY2018-2019-Annual-Technical-Report.pdf

Wu, Y.-C., Lazarus, S. S., & Liu, K. K. (2021). *2017-2018 APR snapshot #24: AA-AAAS participation and performance*. University of Minnesota, National Center on Educational Outcomes. https://nceo.info/Resources/publications/APR_snapshot/brief24

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

Zapata-Rivera, D. (2019). *Score reporting research and applications*. Routledge.

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(4), 442-463. <https://doi.org/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.

Citation:

Clark, A., Kobrin, J. L., Karvonen, M., & Hirt, A. (2023). Teacher use of diagnostic score reports for instructional decision-making in the subsequent academic year. *Practical Assessment, Research, & Evaluation*, 28(6). Available online: <https://scholarworks.umass.edu/pare/vol28/iss1/6/>

Corresponding Author:

Amy Clark
University of Kansas: ATLAS
Email: akclark [at] ku.edu