Chapter 3

Modified Alternate Assessment Participation Screening Consortium: Lessons Learned

Stephen N. Elliott
Arizona State University

Ryan J. Kettler
Rutgers, The State University of New Jersey

Naomi Zigmond
University of Pittsburgh

Alexander Kurz
Arizona State University

This manuscript was supported, in part, by the U.S. Department of Education Office of Elementary and Secondary Education (Grant No. S368A090006). However, the opinions expressed do not necessarily reflect the position or policy of the U.S. Department of Education and no official endorsement should be inferred.

Introduction

The Modified Alternate Assessment Participation Screening (MAAPS) Consortium included the departments of education from Arizona, Pennsylvania, and South Carolina, along with researchers from Arizona State University, Rutgers University, Vanderbilt University, the University of Pittsburgh, and Discovery Education Assessment, with the shared purpose of creating a multi-part screening system for identifying students who would be eligible for an alternate assessment based on modified academic achievement standards (AA-MAS). The MAAPS System included electronic screening tests to predict proficiency in reading and mathematics, as well as a measure of opportunity to learn (OTL) essential academic objectives. The primary goals of the MAAPS Consortium were to (1) develop tools to facilitate educators' accurate assessment for participation decisions for students with disabilities, (2) evaluate the validity and consequences of the participation decision-making tools, (3) apply the MAAPS system for students with disabilities to determine its utility and likely conseqences, and (4) disseminate knowledge learned from the development and implementation of the MAAPS System. The primary outcomes of this project were a new measure of OTL (i.e., *My instructional Learning Opportunity Guidance System or MyiLOGS*) and an objective process for developing a screening assessment to predict end of year test performance.

The MAAPS project started in 2009 and was completed in early 2012. It was designed initially for implementation at the 8th grade level in reading and mathematics and to provide screening data in the form of repeated measures to help educators make decisions with confidence. Related outcomes included examining the relationship between OTL and disability status, sharing information about methods for developing altered items for AA-MAS, and learning about the development of measurement tools that can be extended to other grade levels. The MAAPS Consortium drew from the successful work completed in two completed USDE funded projects: the Consortium for Alternate Assessment Validity and Experimental Studies (CAAVES Project, Compton & Elliott, 2006-2009) and the Consortium for Modified Alternate Assessment Development and Implementation (CMAADI Project, Elliott, Rodriquez, Roach, & Kettler, 2007-2011), as well as on investigators' experience in development and validation of alternate assessments and other educational assessment tools.

Background Context for this Project

The U.S. Department of Education (USDE) cited research findings that suggested there were about 2% of children who were not able to reach grade level standards, even with the best instruction (i.e., Lyon, Fletcher, Fuchs, & Chhabra, 2007). They went on to claim to be building on "what we've learned from science and the field," referencing a review by Lyon et al., (2007) in which the authors noted that even with the best designed instructional interventions, about

1.8% to 2.5% of students were not able to reach grade level standards and those not responding well to the interventions were at risk for later being identified with specific learning disabilities. The USDE also cited work of Torgeson et al. (1999) reporting that all but 24% of the struggling readers who received explicit reading instruction attained average levels (grade level) of achievement, and extrapolated from that finding to speculate that about 2.4% of the total student population might be unable to reach grade level proficiency even with good instruction. They referenced the Lyon et al. (2007) conclusion that when students receive research-based classroom instruction and tutorial interventions, the number of students at risk for learning disabilities is less than 2% of the total student population. Thus, they estimated that 2% of students assessed, or approximately 20% of students with disabilities, was a reasonable and sufficient cap for the new flexibility with AA-MAS.

The USDE also referenced remedial reading research studies of students with reading difficulties in which nearly 80% of the research samples failed to achieve at grade level at the end of the treatment; that extrapolates to 16% of the school population, not 2% of the school population (Francis, Winikates, Mehta, Schatschneider, & Fletcher, 1997; Klingner,Vaughn, Hughes, Schumm, & Elbaum, 1998; Foorman, Francis, Winikates, Mehta, Schatschneider, & Fletcher, 1997). Nevertheless, the new policy was promoted to allow those students with persistent academic disabilities to take an assessment that is more sensitive to measuring their learning progress and that recognizes their individual needs, although they capped the number of scores that could be counted as proficient at 2%.

The USDE spokespersons indicated that IEP teams would likely find it more difficult to identify students eligible for an AA-MAS than for an alternate assessment based on alternate achievement standards (AA-AAS). They suggested that students in the “2% group” would not necessarily be the lowest achieving 2% of students. Nor would the “2% students” be all students with IEPs who are having difficulty with grade-level content or who are receiving instruction below grade level. They challenged states to design criteria to help IEP teams “distinguish between students whose disability has truly precluded them from achieving grade-level proficiency and those who, with appropriate services and interventions, including special education and related services designed to address the student’s individual needs, can be assessed based on grade-level achievement standards” (Federal Register, 2005). Thus, for a student to be eligible for an alternate assessment based on modified academic achievement standards (AA-MAS), she or he must have a disability, there must be evidence that the disability has kept the student from achieving grade level proficiency, and the student’s individualized education program (IEP) must have goals that are based on content standards for the grade in which the student is enrolled. Further, the AA-MAS Non-Regulatory Guidance indicated that:

> *The student’s progress to date in response to appropriate instruction, including special education and related services designed to address the student’s*

individual needs, is such that, even if significant growth occurs, the IEP Team is reasonably certain that the student will not achieve grade-level proficiency within the year covered by the student's IEP. The IEP Team must use multiple valid measures of the student's progress over time in making this determination. (USDOE, 2007, p. 16)

This policy—with its focus on appropriate instruction and the requirement to use multiple valid measures—provided the motivation for the MAAPS project. The desire to compare the predictive validity of two different methods (i.e., tests and teacher ratings) of characterizing student achievement provided additional motivation. Given the struggle with identifying who the 2% eligible students were, the MAAPS team took the perspective that teachers need tools and a systematic way to use them to make decisions during the school year about which students were in need of taking an AA-MAS. The tools needed to efficiently and effectively measure (a) the opportunities students were given to learn the intended curriculum that was also assessed on statewide tests, and (b) how well students were learning the knowledge and skills privileged by the intended curriculum. This information seemed critical to making good eligibility decisions given the AA-MAS participation criteria that were evolving in the partner states in the MAAPS Consortium.

Refining the Measurement of Opportunity-to-Learn

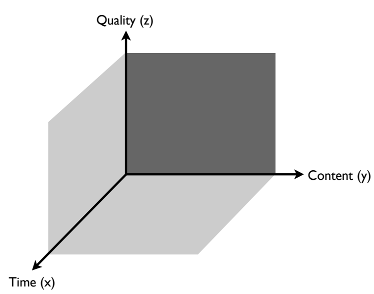
For nearly five decades, researchers have used the concept of opportunity-to-learn (OTL) to examine the inputs and processes necessary for producing important student outcomes. To this end, they have operationalized OTL using various indices along three broad dimensions of the enacted curriculum related to the *time*, *content*, and *quality* of classroom instruction (Kurz, 2011). Anderson (1986) acknowledged the prolific use of the OTL acronym under different conceptual definitions and was one of the first researchers to suggest a merger of the various OTL conceptualizations: "A single conceptualization of opportunity to learn coupled with the inclusion of the variable[s] in classroom instructional research . . . could have a profound effect on our understanding of life in classrooms" (Stevens & Grymes, 1993). Based on a review of the OTL literature, Stevens and Grymes (1993) established the first "unified conceptual framework" of OTL to investigate "students' access to the core curriculum" using four elements: content coverage, content exposure, content emphasis, and quality of instructional delivery.

Despite the fact that Stevens and Grymes did not develop an empirical program of research on the basis of this framework, their conceptualization of OTL has been adopted frequently thereafter (e.g., Abedi, Courtney, Leon, Kao, & Azzam, 2006; Aguirre-Munoz et al., 2006; Herman & Abedi, 2004; Wang, 1998). This "unified" framework, however, fell short of a conceptual synthesis, instead providing three separate "content elements" and one "quality element." In addition, Stevens and Grymes' definitions were too vague to be operational leading researchers to develop a range of disparate indices for each OTL element. Nonetheless, their framework clarified OTL as a teacher effect related to the allocation of adequate instructional time covering a core curriculum via different cognitive demands and instructional practices that produce student achievement.

A Unified Instructional Dimensions OTL Model

The OTL model introduced by Kurz (2011) is situated in the context of the Intended Curriculum Model (ICM) and based on the aforementioned research strands of OTL (see Figure 1). According to Kurz, empirically supported research on OTL at the classroom level has resulted in indicators that fall along three broad instructional dimensions measuring aspects of time, content, and quality that typically co-occur together.

Figure 1. The Unified Instructional Dimensions Model of OTL



Based on this model, students' opportunity to learn the intended curriculum is a matter of degree represented along three orthogonal axes with distinct zero points. Each axis delineates one of the aforementioned instructional dimensions of the enacted curriculum. The model therefore incorporates time-based, content-based, and quality-based OTL conceptualizations as equally valid but limited definitions of OTL that address aspects of the same underlying enacted curriculum. The focus on the enacted curriculum and its temporal, curricular, and qualitative aspects was established on empirical grounds, while the co-occurrence of all three aspects was acknowledged for practical reasons. The conceptual synthesis is further substantiated by a theoretical rationale related to the ICM, which circumscribes the provision of students' opportunity to learn the intended curriculum. (For more information on the ICM, refer to the chapter in this volume on the Consortium for Modified Alternate Assessment Development and Implementation project.)

According to the unified instructional dimensions model of OTL, the first necessary conceptual ingredient of OTL is *time*. To provide students with the opportunity to learn the intended curriculum, teachers must invest instructional time dedicated to the respective knowledge and skills implicated in the intended curriculum. As such, previously used indicators of time such as "allocated time" are not suitable for operationalizing this OTL dimension. Of interest is a teacher's *instructional time* spent on teaching the academic standards of the general curriculum and, if applicable, any intended skills prescribed by a student's IEP. Prior research on time and learning further provides empirical support for examining student engagement and success rate in conjunction with instructional time (e.g., Borg, 1980; Gettinger & Seibert, 2002).

The next instructional dimension that must be integrated into the concept of OTL is *content*. To provide students with the opportunity to learn the intended curriculum, teachers must cover the content implicated in the intended curriculum. Of interest is a teacher's *content coverage* of the academic standards of the general curriculum and, if applicable, any intended objectives prescribed by a student's IEP. Any IEP objectives not directly addressed by the academic standards of the general curriculum are intended to be covered and hence applicable. As such, the "core curriculum" mentioned by Stevens and Grymes (1993) becomes defined in congruence with the legal and legislative mandates of test-based accountability. Previously used OTL indicators related to "tested content" are no longer applicable. As discussed earlier, the normatively desirable target of classroom instruction should be the broader intended curriculum, which subsumes the content of the assessed curriculum.

Only knowing how much time is spent on instruction and what content of the intended curriculum is being covered fails to indicate "how" this time and content were enacted, which requires the integration of a third instructional dimension into the concept of OTL: *quality*. To provide students with the opportunity to learn the intended curriculum, teachers can employ a range of instructional practices that have received empirical support across multiple studies including guided feedback (e.g., Brophy & Good, 1986), reinforcement (e.g., Walberg, 1986), direct instruction (e.g., Gersten, Chard, Jayanthi, Baker, Morphy, & Flojo, 2009), student "think alouds" (e.g., Vaughn, Gersten, & Chard, 2000), and visual representations (e.g., Gersten et al., 2009). In addition, researchers have identified grouping formats other than whole class (e.g., Elbaum, Vaughn, Hughes, Moody, & Schumm, 2000) and cognitive expectations for learning, so-called cognitive demands (e.g., Porter, 2002), as important qualitative aspects of instruction. With respect to cognitive expectations, several classification categories ranging from lower-order to higher-order cognitive processes have been suggested, most notably in Bloom's taxonomy of education objectives (Bloom, 1976). Three quality indicators can be identified: *cognitive expectations*, *evidence-based instructional practices*, and *grouping formats*. A clear theoretical or empirical rationale to preference one indicator over the other is presently not available. All three indicators are therefore presently part of the quality dimension.

The new OTL model further represented each instructional dimension as a continuum that originates in zero. The origin for the *x-axis* indicates that a teacher dedicated zero minutes to teaching the intended curriculum objectives. Conversely, students' opportunity to learn the intended curriculum can be increased by dedicating more instructional minutes to teaching the intended curriculum. Upper constraints are based on allocated time and the total number of school days. Given that the number of school days is very consistent across states (*M* = 180.4, *SD* = .12), the suggested operational index for instructional time (*IT*) is the *average amount of instructional minutes spent on the intended curriculum objectives per day*.

The origin for the *y-axis* indicates that a teacher covered none of the intended curriculum objectives. Students' opportunity to learn the intended curriculum thus can be increased by covering more of the intended curriculum objectives. Upper constraints are based on each state's total number of subject- and grade-specific general curriculum objectives as well as the number of applicable IEP objectives. The suggested operational index for content coverage (*CC*) is the *percentage of addressed intended curriculum objectives*.

The *z-axis* relates to three quality indicators (i.e., cognitive expectations, evidence-based instructional practices, grouping formats). Placing each indicator on a continuum requires a brief explanation. The cognitive process expectations for learning can be grouped along several categories. Although all categories are important, meaningful learning must move beyond expectations of recall/memorization for a transfer of knowledge to occur (see Mayer, 2008). Anderson et al. (2001) further argued:

> When the primary goal of instruction is to promote retention, the focus is on objectives that emphasize *Remember*. When the goal of instruction is to promote transfer, however, the focus shifts to the other five cognitive processes, *Understand* through *Create*. (p. 70)

As such, it seems reasonable to suggest that a teacher's instructional emphasis on high-order/transfer processes can improve the quality of OTL. In addition, the general curriculum standards of virtually all states demand deeper learning beyond recall (e.g., Porter, 2002). The first suggested operational instructional quality index (*CP*) is thus a weighted score that represents the *sum of differentially weighted percentages of instructional time dedicated to each cognitive process expectation*. The two remaining quality indicators can be operationalized in a similar fashion. Teachers are likely to employ a range of generic and evidence-based instructional practices as well as a range of grouping formats from individual to whole class instruction. However, it seems reasonable to argue that teachers who spend more time on evidence-based practices than generic teaching practices improve the quality of students' opportunity to learn the intended curriculum, especially for students with disabilities—likewise for alternative grouping formats. As such, the second suggested operational quality index (*IP*) is the *sum of differentially weighted percentages of instructional time dedicated to each instructional practice*. Similarly, the third suggested operational quality index (*GF*) is the *sum of differentially weighted percentages of instructional time dedicated to each grouping format*. These weighted scores—*CP*, *IP*, *GF*—and their specific weights will be further operationalized based on the methodological conventions of the OTL measure. More details are provided in the section on the specific MyiLOGS OTL Indices (p. 12).

The origin for the *z-axis* thus indicates that no teaching occurred at all. Whenever a teacher spends time on instruction, he or she must place instructional emphases along different cognitive expectations, instructional practices, and grouping formats. As such, instructional quality can only range from low to high, depending on which type of expectations (low-order vs. high-order), practices (generic vs. evidence-based), and formats (alternative vs. whole class) were emphasized. Table 1 summarizes the instructional dimensions of the proposed OTL model and its respective indicators, definitions, and suggested operational indices.

Table 1. Instructional Dimensions, Indicators, Definitions, and Operational Indices of OTL

Dimension	Indicator	Definition	Index
Time	Instructional Time	Instructional time dedicated to teaching the general curriculum standards and, if applicable, any intended IEP objectives.	***IT***: Average amount of instructional minutes spent on intended curriculum objectives per day.
Content	Content Coverage	Content coverage of the general curriculum standards and, if applicable, any intended IEP objectives.	***CC***: Percentage of addressed intended curriculum objectives.
Quality	Cognitive Processes	Emphasis of cognitive process expectations along a range of lower-order to higher-order thinking skills.	***CP***: Sum of differentially weighted percentages of instructional time dedicated to each cognitive process expectation.
	Instructional Practices	Emphasis of instructional practices along a range of generic to empirically supported practices.	***IP***: Sum of differentially weighted percentages of instructional time dedicated to each instructional practice.
	Grouping Formats	Emphasis of grouping formats along a range from individual to whole class instruction.	***GF***: Sum of differentially weighted percentages of instructional time dedicated to each grouping format.

Note: Emphasis can be operationalized as the amount of instructional minutes.

In summary, the conceptual synthesis of OTL has resulted in defining students' opportunity to learn the intended curriculum on the basis of three empirically supported instructional dimensions: time, content, and quality. On the basis of theory and research, we established OTL indicators and provided suggestions for operationally defined indices. This integrated concept of OTL is consistent with the legal and legislative demands of test-based accountability and builds upon previous curriculum and OTL frameworks. As such, we defined OTL for purposes of the MAAPS Project as *the degree to which a teacher dedicates instructional time and content coverage to the intended curriculum objectives emphasizing high-order cognitive processes, evidence-based practices, and alternative grouping formats*. This definition was used to guide the development of a new measure of OTL that teachers could use to document instruction for an entire class, as well as for individual students with disabilities.

MyiLOGS: A New Measure of OTL

This online technology (www.myilogs.com) is designed to assist teachers with the planning and implementation of intended curricula at the class and student level. MyiLOGS was developed on the theoretical and empirical basis of the OTL research literature including the previously discussed curriculum framework of the ICM and the conceptually integrated model of OTL. As such, this educational technology can be used to document all three instructional dimensions of the enacted curriculum via indicators of instructional time, content coverage, and instructional quality such as cognitive process expectations, instructional practices, and grouping formats.

MyiLOGS features the state-specific academic standards of the general curriculum for various subjects and additional customizable skills that allow teachers to add student-specific objectives (e.g., IEP objectives). The tool therefore allows teachers to document the extent to which their classroom instruction covers individualized intended curricula. To this end, MyiLOGS provides teachers with a monthly instructional calendar that includes an expandable sidebar, which lists all intended objectives for a class. Teachers drag-and-drop planned skills that are to be the focus of the lesson onto the respective calendar days and indicate the approximate number of minutes dedicated to each skill. After the lesson, teachers are required to confirm enacted skills, instructional time dedicated to each skill, and any time not available for instruction (due to transitions, class announcements, etc.) at the class level. In addition, two randomly selected days per week require further documentation. On these sample days, teachers report on additional time emphases related to the skills listed on the calendar including cognitive expectations, instructional practices, grouping formats, engagement, goal attainment, and time not available for instruction. This detailed reporting occurs at the class and student level along two two-dimensional matrices and two ratings. Teachers can further review a range of charts and tables that provide detailed information on their enacted curriculum and its relation to the intended curriculum (i.e., subject-specific academic standards and custom objectives). These instructional reports are available for the entire class and individual students. However, this functionality was not available to teachers during the course of this study. Screenshots of the

MyiLOGS calendar interface as well as the sample day matrices and ratings are displayed in Figures 2 and 3.

For the first matrix, teachers report on the instructional minutes allocated per skill along five cognitive process expectations for student learning adapted from the revised version of Bloom's taxonomy (see Anderson et al., 2001). For the second matrix, teachers report on the instructional minutes allocated per instructional practice along three grouping formats. Teachers further rate engagement and goal attainment along a 4-point scale. Student engagement and successful work completion are two previously discussed indicators for purposes of determining academic learning time. The definitions for the cognitive process expectations and instructional practices are provided in Tables 2 and 3. The grouping formats were defined as follows: (a) *Individual*: Instructional action is focused on a single individual; (b) *Small group*: Instructional action is focused on a small groups; (c) *Whole Class*: Instructional action is focused on the whole class.

Figure 2. Screenshot of the MyiLOGS Calendar Interface

Figure 3. MyiLOGS Sample Day Matrices and Ratings

Estimated Time Allocation Across Cognitive Process Dimensions for: Tunnell Gr. 8 Math

Skill	Attend	Remember	Understand/Apply	Analyze/Evaluate	Create	Sum	Calendar Minutes
S3C3PO3 Linear equations and inequalities	5	20	45	0	0	70	70
Concept Review Bell Work	0	5	5	0	0	10	10
Time Not Available for Instruction						0	0
					Update Totals Total:	80	80

Estimated Time Allocation Across Instructional Practices for: Tunnell Gr. 8 Math

Teacher Actions	Individual	Small Group	Whole Class	Sum
Provided Direct Instruction	0	0	10	10
Provided Visual Representations	0	0	5	5
Asked Questions	0	5	5	10
Elicited Think Aloud	0	0	0	0
Used Independent Practice	0	0	0	0
Provided Guided Feedback	5	5	0	10
Provided Reinforcement	0	0	10	10
Assessed Student Knowledge	0	0	0	0
Other Instructional Practices	0	0	35	35
Time Not Available				0
		Update Totals	Calendar Total: 80	80

Engagement Matrix for: Tunnell Gr. 8 Math

Class Engagement	Learning Goal Attainment
Not Engaged (0%)	No effort or product observed (0%)
Low % of time (<50%)	Low effort or limited portion of work completed (<50%)
Moderate % of time (50% - 80%)	Moderate effort or moderate portion of work completed (50% - 80%)
High % of time (>80%)	High effort or substantial portion of work completed (>80%)

Table 2. Cognitive Process Expectations for Student Learning and Definitions

Cognitive Process	Definition
Attend	Orient toward instructional task and related instructions. • Synonyms include *listen, focus, pay attention*.
Remember[a]	Retrieve relevant knowledge from long-term memory. • Synonyms include *recognize, identify, recall, retrieve*.
Understand[a]	Construct meaning from instructional messages. • Synonyms include *interpret, exemplify, classify, summarize, infer, compare, explain*.
Apply[a]	Carry out or use a procedure in a given situation. • Synonyms include *execute, implement, use*.
Analyze[a]	Break materials into its constituent parts and determine how the parts relate. • Synonyms include *differentiate, organize, integrate, attribute*.
Evaluate[a]	Make judgments based on criteria and standards. • Synonyms include *check, test, critique, judge*.
Create[a]	Put elements together to form a coherent whole or a new structure. • Synonyms include *generate, hypothesize, plan, design, produce*.

[a]This cognitive process definition is based on the revised Bloom's taxonomy (see Anderson et al., 2001)

To minimize teachers' response burden for purposes of this study, the related cognitive processes *Understand* and *Apply* as well as *Analyze* and *Evaluate* were collapsed in the cognitive process matrix. The relation and grouping of these cognitive processes is supported by Webb's DOK levels: (a) the learning expectations under Understand/Apply are mostly limited to routine applications of comprehension and execution linked to familiar skills and concepts; and (b) the learning expectations under Analyze/Evaluate mark a shift toward more complex thinking that requires abstract reasoning, planning, developing, and using of evidence (Webb, 2006). In this study, the cognitive process matrix further included the *Attend* category, which is not part of the revised Bloom's taxonomy (see Anderson et al., 2001). The cognitive expectation of *Attend* allowed teachers to differentiate between the expectation of students (passively) listening to instructional tasks and related instructions and (actively) recalling information such as a fact, definition, term, or simple procedure. A similar category of cognitive demand has been used previously in the context of special education, especially for students with significant cognitive disabilities (Karvonen, Wakeman, Flower, & Browder, 2007).

Table 3. Instructional Practices and Definitions

Instructional Practice	Definition
Provided Direct Instruction[a]	Teacher presents issue, discusses or models a solution approach, and engages students with approach in similar context.
Provided Visual Representations[a]	Teacher uses visual representations to organize information, communicate attributes, and explain relationships.
Asked Questions[a]	Teacher asks questions to engage students and focus attention on important information.
Elicited Think Aloud[a]	Teacher prompts students to think aloud about their approach to solving a problem.
Used Independent Practice	Teacher allows students to work independently to develop and refine knowledge and skills.
Provided Guided Feedback[a]	Teacher provides feedback to students on work quality, missing elements, and observed strengths.
Provided Reinforcement[a]	Teacher provides reinforcement contingent on previously established expectations for effort and/or work performance.
Assessed Student Knowledge	Teacher uses quizzes, tests, student products, or other forms of assessment to determine student knowledge.
Other Instructional Practices	Any other instructional practices not captured by the aforementioned key instructional practices.

[a]This instructional practice has received empirical support across multiple studies.

The second matrix lists nine instructional practices and three grouping formats. In Table 3, nine instructional practices are marked by a table note to indicate empirical support on the basis of research syntheses and meta-analyses (e.g., Brophy & Good, 1986; Gersten et al., 2009; Marzano, 2000; Swanson, 2000; Vaughn et al., 2000; Walberg, 1986). In addition, grouping formats other than whole class also have received empirical support for improving learning outcomes (see Elbaum et al., 2000). “Other instructional practices” represents a “catch-all” category to allow teachers to report on their entire allocated time per class using the available selection of instructional practices or “time not available for instruction.” Teachers use the latter category to indicate any non-instructional minutes (e.g., transitions, announcements, fire drills), which together with instructional minutes should add up to the total allocated class time (e.g., 90-minute ELA class).

MyiLOGS OTL Indices. Based on our definition of OTL, the instructional data collected via MyiLOGS were used to derive several OTL indices. First, the instructional time index was specified into three separate indices: (a) instructional time spent on state-specific academic standards (*Time on Standards*); (b) instructional time spent on custom objectives (*Time on Custom*); and (c) non-instructional time (*Non-Instructional Time*). These time-based indices were calculated based on average minutes per day and as average percentages of allocated class time. The latter convention was used to allow for comparability between classes that differed in allocated class time. Second, the content coverage index (*Content Coverage*) was based on the percentage of state-specific academic standards a teacher addressed for at least one minute or more throughout the entire logging period. Lastly, all time-based and content-based OTL indices were calculated on the basis of calendar days and detail days with the former representing the largest set of data points.

Quality-related indices were based on instructional time emphases allocated to the various cognitive processes (CP), instructional practices (IP), and grouping formats (GF). Given the focus on high-order thinking skills, evidence-based instructional practices, and grouping formats other than whole class, summary scores were calculated for CP, IP, and GF reflective of the respective emphases. First, instructional time allocations across the various CP, IP, and GF categories were converted into percentages. Second, a weight of 1 was applied to all lower-order thinking skills, generic instructional practices, and whole class instruction for CP, IP, and GF percentages, respectively. A weight of 2 was applied to all high-order thinking skills, empirically supported practices, and individual/small group instruction for CP, IP, and GF percentages, respectively. As such, all cognitive expectations, instructional practices, and grouping formats received credit; yet those presumed to contribute more to enhance the quality of OTL received a greater weight. The CP, IP, and GF scores thus ranged between 1.00 and 2.00. A CP, IP, and GF score of 1.00 indicates an exclusive focus on lower-order thinking skills (i.e., attend, remember), generic instructional practices (i.e., independent practice, other instructional practices), and whole class instruction, respectively. A CP, IP, and GF score of 2.00, on the other hand, indicates an exclusive focus on higher-order thinking skills (i.e., understand/apply, analyze/evaluate, create), evidence-based instructional practices (i.e., direct instruction, visual representations, questions, think aloud, guided feedback, reinforcement, assessment), and individual/small group instruction, respectively. The three quality-based OTL indices were calculated on the basis of detail days only.

Using MyiLOGS with Integrity and Reliability: Competency-Based Training. To ensure that all users of MyiLOGS are familiar with the coding conventions and accurate use of the system (e.g., cognitive process expectations, differentiated instruction), each teacher had to participate in a professional development training. The first introductory element is centered around a video supported *worked example* lasting about 30 minutes, which provides a step-by-step demonstration of how to complete the three essential MyiLOGS tasks (i.e., daily calendar, sampled class details, sampled student details). The second element is a *guided practice* session lasting about

2 hours. During that time, a trainer models the steps for completing each task followed by teachers practicing these steps with the support of each other and additional trainers. To establish the definitions of the cognitive process expectations and instructional practices, teachers complete worksheets that ask them to define each category in their own words and provide examples. Subsequent discussion and modeling is used to resolve any questions and disagreements. The third element features the MyiLOGS *performance assessment* lasting about 1 hour. To ensure teachers have mastered the logging conventions of the technology to accurately represent their instruction (e.g., differentiated instruction, substitute instruction, student absences), teachers have to pass a sequence of performance tests. These tests feature written instructional scenarios that summarize typical lessons along the calendar, class, and student level. Figure 4 displays an example of an instructional scenario. Teachers have to correctly log the instructional scenario via the MyiLOGS software. Once completed, a trainer reviews the accuracy of the logged scenario. Teachers have to pass two scenarios with 100% accuracy to be able to continue in the study. A total of five independent instructional scenarios were available to teachers in the allotted training time.

Figure 4. Instructional Scenario Example Used in the MyiLOGS Performance Assessment

SUBJECT: Mathematics (Gr. 8)
CLASS PERIOD: **60 min**

CALENDAR:

- ***Simplify numeric expressions*** for about 60 minutes. **[Numbers and Operations]**
- James showed up 10 minutes late (***time not available for instruction***).

CLASS ENACTED:

- For review, you **asked questions** of the **whole class**. Students were expected to **recognize** some of the basic rules for ***simplifying numeric expressions*** (e.g., order of operations) for about 10 minutes.
- For the remaining 50 minutes, you **modeled and discussed** several problems and their solutions on the board for the **whole class**. During that time, you provided **guided feedback** to the **whole class** (about 10 minutes) and some **positive reinforcement** to **individual students** (about 5 minutes). Students were expected to **attend** to your models and explanations (about 20 minutes) and **recognize** what strategy you just applied (about 30 minutes).
- Overall, the class was **highly engaged** and completed **all required work**.

STUDENT ENACTED:

- **Kayla** participated and completed the same activities as the rest of the class. Her engagement and effort were moderate today.
- **James** showed up 10 minutes late and thus missed the review. Otherwise, he participated and completed the same activities as the rest of the class. His engagement and effort were low today.

The fourth element is an *independent practice* session lasting about 1 hour. During that time, teachers are allowed to use their teaching materials such as lesson plans and textbooks to retrospectively log the previous month of instruction at the calendar level.

Classroom Observation System to Estimate Validity of MyiLOGS Teacher-Reported Data. To estimate the extent to which teachers were using MyiLOGS reliably, each teacher participant was observed at least once during his or her login period. In addition, a subset of three teachers per state was randomly chosen for two additional observation sessions to determine the stability of the reliability estimates. To this end, we developed an observation form that mirrored the two two-dimensional matrices displayed in Figure 5. Trained observers used this form to code the dominant cognitive expectation for student learning and instructional practice observed during a 1-minute interval. A vibrating timer on a fixed interval was used to indicate the 1-minute recording mark.

For observation purposes, all classrooms observers (a) prerecorded the skills listed on the MyiLOGS calendar for the given day; (b) started the 1-minute interval with the bell or at the lesson's designated start time; (c) made a tally in both matrices according to the cognitive expectation and instructional practice that occupied the majority of the time during a 1-minute interval (by skill and grouping format); and (d) kept a frequency count of discreet events such as brief praise statements. At the conclusion of the observation, the observer was allowed to make time adjustments to reflect the summative duration of discreet events as well as the MyiLOGS convention of equal emphasis. The latter convention requires teachers to divide instructional minutes equally according to emphasis. For example, a teacher who allowed students to work independently for 10 minutes but concurrently provided students with individual guided feedback throughout the entire time could not log 10 minutes under each practice. Instead, the teacher must divide the instructional minutes accordingly (i.e., 5 minutes per practice). This convention constrains teachers to the allocated class time—the more skills or practices that are addressed, the less instructional time can be dedicated to each one. Accordingly, observers were allowed to make tally adjustment immediately following the observation.

For agreement purposes, cell-by-cell agreement was calculated for each matrix based on cell estimates within a 3-minute range or less. That is, two observer estimates of direct instruction at the whole class level of 20 minutes and 23 minutes respectively were counted as an agreement. Likewise, teacher and observer estimates of the Pythagorean Theorem at the Remember level of 4 minutes and 0 minutes respectively were counted as a disagreement. For each matrix, interrater agreement was calculated as the total number of agreements divided by the sum of agreements and disagreements. In addition, a combined interrater agreement percentage was calculated as the total number of agreements across both matrices divided by the sum of agreements and disagreements across both matrices. That latter index was used in establishing the training criterion ($\geq$ 80%) and retraining criterion (< 80%) for observers.

Figure 5. MyiLOGS Classroom Observation Form

Teacher ID:	Date:	Class:	Time:

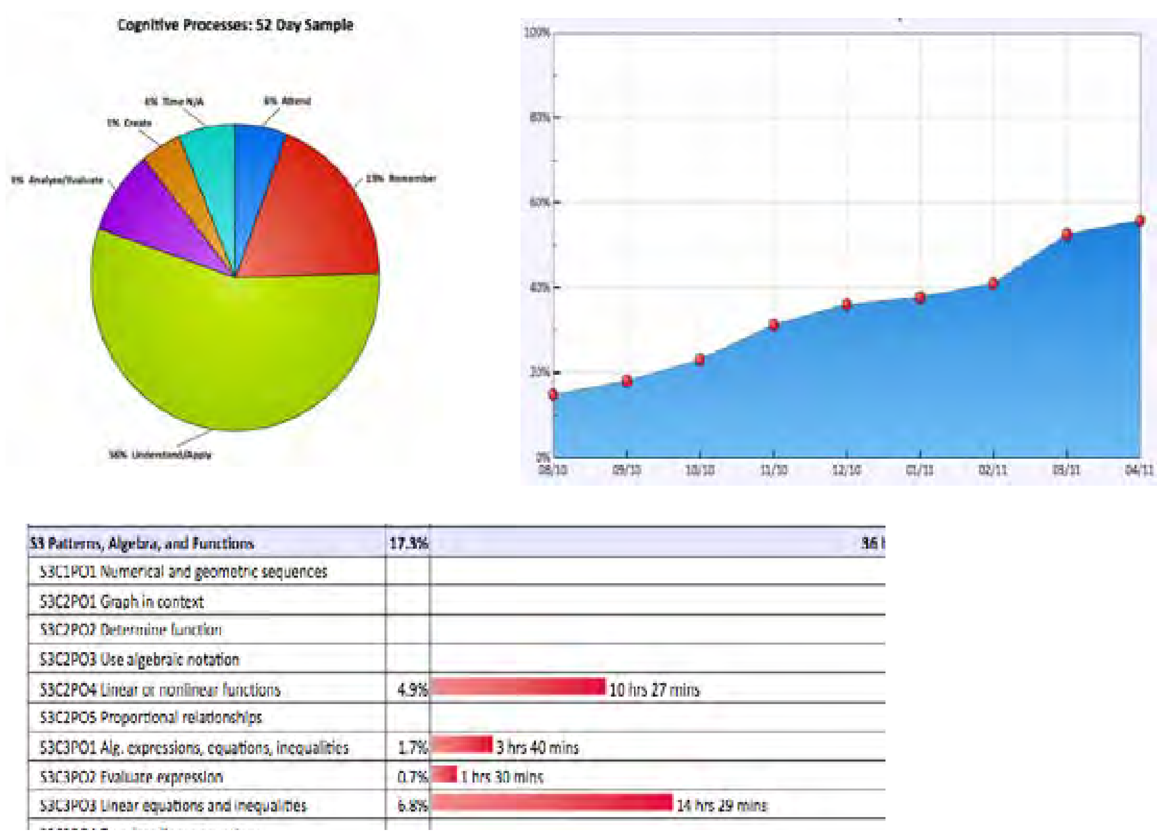
① Record in 1-min intervals. ② Use tally marks (| || ||| ~~||||~~) to record the **student expectation** and **teacher action** that occupied the **majority of time** during the 1-min interval.

Skills	Attend Listen, focus, pay attention	Remember Recognize, identify, recall	Understand/Apply Interpret, exemplify, classify, summarize, infer, compare, explain / Execute, implement, use	Analyze/Evaluate Differentiate, organize, integrate, attribute / Check, test, critique, judge	Create Generate, hypothesize, plan, design, produce
Time not available for instruction					

Teacher Actions	Individual Action is focused on single individuals	Small Group Action is focused on small groups	Whole Class Action is focused on entire class
Provided Direct Instruction Teacher presents issue, discusses or models a solution approach, and engages students with approach in similar context.			
Provided Visual Representations Teacher uses visual representations to organize information, communicate attributes, and explain relationships.			
Asked Questions Teacher asks questions to engage students and focus attention on important information.			
Elicited Think Aloud Teacher prompts students to think aloud about their approach to solving a problem.			
Used Independent Practice Teacher allows students to work independently to develop and refine knowledge and skills.			
Provided Guided Feedback Teacher provides feedback to students on work quality, missing elements, and observed strengths.			
Provided Reinforcement Teacher provides reinforcement contingent on previously established expectations for effort and/or work performance.			
Assessed Student Knowledge Teacher uses quizzes, tests, student products, or other forms of assessment to determine student knowledge.			
Other Instructional Practices Any instructional practices not captured by the aforementioned key instructional practices. You can use the class notes to leave additional details.			
Time not available for instruction			

Instructional Feedback Reports for Teachers. Users of MyiLOGS receive instructional feedback reports on demand, although teachers in the MAAPS project did not get reports until several months after their data collection was completed because the reporting format needed to be developed with insightful users. Future teachers can review their MyiLOGS instructional feedback reports after logging approximately one month of instruction. These reports include tables and figures that detail a teacher's instructional provisions on the basis of the various OTL indices collected via MyiLOGS. Figure 6 shows a collection of three charts: (a) the pie chart provides a breakdown of the various instructional practices emphasized based on a teacher's time allocations; (b) the area chart displays a teacher's cumulative content coverage of the state standards across the school year; and (c) the bar chart details a teacher's time allocations for the various state standards by the respective domain. Over a dozen charts are available to teachers, several of which provide information for the overall class as well as individual students. Teachers thus have the ability to not only monitor the instructional provisions for their overall class but also the extent to which they differentiate their instruction for specific students.

Figure 6. Examples from the MyiLOGS Instructional Feedback Reports



Evidence to Support Reliability and Validity of MyiLOGS

We examined the usability and initial validity evidence for MyiLOGS with a three-state sample of 38 middle school teachers and found they could use the tool with high integrity and moderately high reliability during an average of 151 school days. In addition, from a number of sources of validity evidence from teachers' MyiLOGS database there was empirical support for the following findings: (a) teachers could be trained to report reliably on various OTL indices that provide a valid account of classroom instruction as supported by third party observations; (b) the applied online technology offered teachers a feasible and time efficient method for collecting OTL data at the class and student level on a daily basis across the school year; (c) the resulting system shows promise for a large-scale collection of OTL data because of its technical qualities and high acceptance by users. A detailed report of the data behind this summary of technical qualities of MyiLOGS has been written by Kurz, Elliott, and Kettler (2012) and available at the MyiLOGS website. Future OTL research is needed to confirm these findings due to limitations based on our relatively small and volunteer sample.

Results of the MAAPS OTL Study

Three major categories were implicit in the data set: (a) state (i.e., Arizona, Pennsylvania, and South Carolina); (b) subject (i.e., MA and ELA); and (c) class type (i.e., general education class and special education class). Arizona represented a unique sample, because all class types in this subsample were general education classrooms. As such, Arizona represents the full inclusion model, whereas the other two states featured a mix of full-inclusion general education classrooms and special education classroom. However, given the inclusion of all target students in the regular state assessment, the instructional provision of the general curriculum standards was fully warranted for both class types across states. That is, all students in the respective classes should have had the opportunity to learn the academic standards of the general curriculum (which were subsequently assessed via the respective state test) and any other IEP mandated objectives. At the time of the study, the state of Arizona mandated teachers cover 100% of the general curriculum standards.

With respect to basic time and content frameworks, teachers within and between states demonstrated a great deal of variation both in terms of allocated class time and the number of academic standards for each subject area. Across states and subject areas, the allocated class time ranged between 25 and 150 minutes and the number of academic standards ranged between 32 and 115. Variability in time extended further including for teachers of the same subject in the same state: (a) allocated class time in MA ranged between 46-120 minutes, 39-82 minutes, and 30-70 minutes in Arizona, Pennsylvania, and South Carolina, respectively; and (b) allocated class time in ELA ranged between 57-150 minutes, 39-82 minutes, and 25-70 minutes in Arizona, Pennsylvania, and South Carolina, respectively. Within these basic frameworks of allocated class time and number of content standards, teachers further varied in the extent to which they dedicated instructional time to the content standards and different custom skills, as well as the extent to which allocated time was non-instructional (e.g., transitions, announcements). Irrespective of the large standard deviations, the average percentage-based indices across states were similar for MA and ELA with 69% and 66% for *Time on Standards*, 27% and 28% for *Time on Custom*, 4% and 5% for *Non-Instructional Time*, as well as 66% and 69% for *Content Coverage*, respectively.

The extent to which the observed variation and values were a function of class type was also examined by considering general and special education classes across states separately. The range in allocated class time remained wide for both class types with 39-150 minutes in general education classes (range = 111 minutes) and 25-82 minutes in special education classes (range = 57 minutes). The variation around the percentage-based time and content indices was greater for special education classrooms than general education classroom. On average, the percentage of instructional time dedicated to the standards was greater in general education classrooms (71%) than in special education classrooms (61%). On the other hand, the average percentage of instructional time dedicated to custom skills (e.g., IEP objectives) was greater in special education classrooms (30%) than in general education classrooms. The average percentage of non-instructional time was similar in both class types. Lastly, the average percentage of content coverage was greater in general education classrooms (74%) than in special education classrooms (54%). The differences in percentage-based indices for *Time on Standards* and *Content Coverage* further exhibited medium effect sizes.

Assuming that academic achievement is higher in general education classrooms, the findings that general education teachers were able to dedicate more instructional time to teaching the academic standards and cover more content standards were not surprising. However, students in this study's special education classrooms nonetheless participated in the same regular state assessments as their general education peers, which should have necessitated the same academic expectations for both subgroups irrespective of instructional setting. In fact, it seems reasonable to suggest that students' placement in special education due to disability-related academic difficulties should result in even greater time and content emphasis on the academic standards of the general curriculum precisely because of their disability-related academic challenges (e.g., attention difficulties, memory issues, behavioral challenges). The present results for this sample, however, do not support the notion of *equal* OTL for students with disabilities based on class type.

With respect to OTL indices for instructional quality, data were collected on two random days per week. That is, teachers completed additional information on cognitive processes, instructional practices, grouping formats, class engagement, and goal attainment/effort. Specifically, teachers logged quality-related OTL indices for an average of about 43 school days, or 24% of the school year. Based on summary data across states, subject-specific differences in OTL indices were noted along the *Cognitive Process*, *Instructional Practice*, and *Grouping Format* scores. These summary indices indicated a greater emphasis of high-order thinking skills in ELA than in MA, a greater emphasis of evidence-based practices in MA than in ELA, and a greater emphasis of alternative grouping formats in MA than in ELA. None of these general trends, however, represented statistically significant differences based on this sample.

Subsequent descriptions of total time allocations across the different cognitive process, instructional practices, and grouping formats indicated the following. Across states, the most emphasized cognitive processes were *Understand/Apply*. The *Remember* process was more prevalent in MA than in ELA, and the *Create* process more prevalent in ELA than in MA. Both findings appear reasonable given the large number of memorable MA facts and the ability for ELA teachers to utilize the *Create* process during composition tasks. With respect to instructional practices, *Independent Practice* represented the most commonly emphasized practice among available choice across both subject areas. Moreover, *Direct Instruction* and *Assessed Student Knowledge* followed *Independent Practice* as the second and third order of emphasis across subject areas. Lastly, *Whole Class* was the most commonly emphasized grouping format across subject areas. Conversely, *Small Group* represented the least commonly emphasized grouping format across subjects.

In the context of class type, differences in quality-related OTL scores were statistically significant for both the *Cognitive Process* and the *Grouping Format* scores with large effect sizes. That is, students in general education classrooms experienced a greater emphasis of high-order cognitive processes and a greater emphasis of whole class instruction than students in special education classrooms. An examination of the total time allocations indicated that the major difference in cognitive processes between both class types was largely due to a greater emphasis of *Attend* in special education classrooms with a large effect size and a greater emphasis of *Analyze/Evaluate* in general education also with a large effect size. With respect to instructional practices, students in general education classrooms experienced a greater emphasis on *Assessed Student Knowledge* with a large effect size. In addition, it should be noted that *Independent Practice* remained the most emphasized instructional practice in both classroom settings. Not surprisingly, the major difference in grouping formats between both class types was due to a significantly greater emphasis of *Individual* grouping formats in special education classrooms and a significantly greater emphasis of *Whole Class* grouping formats in general education classrooms.

To examine the extent to which teachers provided a differentiated opportunity structure for students with disabilities compared to their peers in the same class, teachers were asked to report on sample-day details at the class and student level. On average, teachers logged about 43 sample days, or 24% of the school year. A comparison of the class-based and student-based OTL indices across subject areas and states indicated five statistically significant differences, three of which yielded effect sizes above .20. Compared to the overall class, students with disabilities experienced less *Time on Standards*, more *Non-Instructional Time*, and less *Content Coverage* than their classmates. Statistically significant difference for two OTL indices related to instructional quality, the *Cognitive Process Score* and the *Instructional Practice Score*, were also found. However, the effect sizes for both indices were very small. These results were based on summary data across states, subject areas, and class types.

Looking at individual states, the results based on the Pennsylvania subsample differed from the remaining two states. In Pennsylvania, only two indices, *Time on Standards* and *Content Coverage*, showed statistically significant differences between the class and student level; however, the magnitude of the difference was very small. The largest differences were found in the Arizona subsample, where six of the seven OTL indices showed statistically significant differences between the class and student level. In terms of effect size, the results indicated medium effect sizes for *Time on Standards*, *Non-Instructional Time*, and *Content Coverage*. The fact that the Arizona subsample was comprised exclusively of general education classes presents a possible explanation for the larger effect sizes. That is, the Arizona subsample represented the full inclusion model, where students with disabilities are included in a class of general education peers who are likely to perform at higher academic levels. Consequently, teachers may be able to provide more instructional time on standards-based instruction to students who are academically ready to benefit, namely the majority of classmates without disabilities. However, it should be noted that students with disabilities did not receive significantly different time allocations to *Time on Custom* skills/activities compared to their overall class; a category reserved for any academic objectives or activities that are not part of the general curriculum standards. In fact, a review of the 554 custom skills/activities logged in all 46 classrooms indicated that only 1 custom skill/activity was tagged as an IEP objective related to reading fluency. Furthermore, over 50% of custom skills logged were based on summary activities that either practiced or reviewed standards-related instruction such as "Bell Work" or "Review," as well as technology-based activities such as Study Island or ALEKS®.

The issue of *Non-Instructional Time* also warrants additional consideration. With the exception of the Pennsylvania subsample, target students (with disabilities) experienced more *Non-Instructional Time* than their classmates. The *Non-Instructional Time* index is intended to reflect any teacher-reported minutes of allocated class time that could not be used for instruction (either on general curriculum standards or custom skills/activities). However, teachers were not asked to identify the types of non-instructional activities such as transitions, school announcements, and so on. The magnitude of the difference between the class and student level was the largest in the Arizona subsample, where teachers provided data on OTL for target students (with disabilities) and the overall class (largely without disabilities). The reasons why these students with disabilities experienced more *Non-Instructional Time* than their classmates, however, remain unclear (e.g., behavioral challenges, absences, related services provisions).

A comparison of the differentiated opportunity structure by classroom type indicated that in special education classes the differences in OTL indices between the class and student level were statistically significant, albeit with very small effect sizes for *Time on Standards* and *Content Coverage*. In contrast, six of the seven OTL indices in general education classrooms showed statistically significant differences with a range of small and medium effect sizes. Specifically, the magnitude of the difference for *Time on Standards*, *Non-Instructional Time*, and *Content Coverage* yielded effect sizes above .20. A comparison of the findings by class type thus indicated that the differences in OTL indices were largely a function of class type. The gap in OTL for *Instructional Time* between the class and student level was larger in general education classes (.24) than in special education class (.18). Moreover, the gap in OTL for Content Coverage was comparatively small (.08) between the class and student level in special education classes compared to general education classes (.31).

In summary, the findings support the contention that OTL is a differentiated opportunity structure, which differs at the class and student level. However, it should be noted that in this study the student level was comprised of students with disabilities of low academic performance. Second, the differences in OTL indices were largely related to class type, with general education classes yielding the largest OTL gaps for students with disabilities. That is, students with disabilities in this project who were taught in general education classes experienced (a) less instructional time on state-specific standards than their classmates; (b) more non-instructional time than their classmates; and (c) less content coverage of the states-specific standards than their classmates. *These results extend the findings of the previous research question, which already indicated unequal OTL* between *different class types. The findings of this question provided further evidence of unequal OTL* within *class types.*

Lessons Learned about OTL from MyiLOGS Research

The iterative development process of MyiLOGS throughout the MAAPS project has resulted in an OTL measurement tool that can be applied at scale on a daily basis in a way that is technically adequate as well as usable, feasible, and promising (Kurz, Elliott, & Kettler, 2012). By incorporating teacher feedback via pilot testing, we were able to create an efficient online teacher log that teachers were able to complete several times a week with an average time investment of 6 minutes per week. Consequently, we were able to ascertain a comprehensive data set on daily classroom instruction that was comprised, on average, of about 150 school days. Moreover, a robust PD workshop allowed virtually all participants to be trained to criterion. Classroom observations further confirmed that teachers were able to log their instruction comparable to an independent observer logging the same lesson. The following lessons learned can be summarized:

- OTL can be operationalized along three enacted curriculum dimensions—time, content, and quality—to create a unified model.
- MyiLOGS represents an example of how indices of instructional time, content, and quality can be efficiently measured in an online format that results in reliable indices and valid inferences about teachers' instruction.
- MyiLOGS represents an example of how instructional feedback reports based on teachers' OTL indices can provide meaningful information for instructional changes (as reported by teacher surveys).
- The data from this initial project indicated that OTL varies between the overall class and individual students with disabilities. Moreover, students with disabilities nested in general education classes experience less instructional time on standards, more non-instructional time, and less content coverage.

The data from this initial project also indicated that some OTL indices are related to students' achievement, underscoring the importance of interpreting student outcomes data in the context of instructional data on inputs and processes.

Developing a Proficiency Screening Measure

Possible predictors of future proficiency on large scale achievement tests include previous performance on similar tests, performance on screening tests, and structured teacher ratings of achievement. While the former has the advantages of being readily available in many cases and sharing a common methodology, screening tests and teacher ratings are time efficient methods for obtaining information where previous achievement scores are missing, or for adding information and measures to meet the "reasonably certain" and "multiple valid measures" criteria included in the policy. Previous research indicates that both computer-based multiple-choice tests and the Performance Screening Guides (PSGs) of the Social Skills Improvement System (SSiS; Elliott & Gresham, 2007) can be accurate predictors of performance on future achievement proficiency tests (Kettler, 2011; Kettler, Elliott, Davies, & Griffin, 2012). The purpose of the study was to determine which of these scores or combinations of scores worked best for making such predictions. (The terminology used herein—specifically including the terms predicting, prediction, and predictor—aligns with current measurement theory and does not necessarily represent specific policies of partner states in the current project.)

Developing the Screening Tests

New screening tests were developed using systematic design based on test theory and pilot testing. A team of researchers and content experts from the three MAAPS states developed multiple-choice, online screening tests of reading and mathematics achievement at the eighth grade level. These tests were developed from items selected from an original pool of over 100 in reading and 100 in mathematics, provided by Discovery Education Assessment (DEA). The items on each test represented key subdomains in each content area. The states varied greatly in their conceptualization of reading or language arts subdomains. All three states included items indicative of vocabulary, comprehension, and interpretation, so these three subdomains were used to categorize items on the MAAPS reading screening test. Subdomains in mathematics were highly consistent, including numbers, geometry, measurement, algebra, and data analysis. Table 3 indicates the percentage of items dedicated to each subdomain on achievement proficiency, based on each state's eighth grade test blueprint, as well as on the MAAPS screening tests.

Two forms of the reading test and one form of the mathematics test were completed by small samples of 8th grade students in Arizona and Pennsylvania. A reading form with 26 items had a Cronbach's alpha of .86 (n = 67), and another with 30 items had a Cronbach's alpha of .90 (n = 63). The mathematics form with 27 items had a Cronbach's alpha of .79 (n = 44). By design, all of these forms were longer than the final versions of the screening tests. Items were selected for the final versions of the reading and mathematics screening tests based on difficulty, item-to-total correlation, and content coverage. The final version of the reading screening test included 22 items, with 6 on the Vocabulary subscale, 8 on the Comprehension subscale, and 8 on the

Interpretation subscale. The final version of the mathematics screening test included 26 items, with 10 on the Numbers subscale, which was heavily weighted because it is a prerequisite for all of the other areas. The mathematics screening test also included 3 items on the Measurement subscale, 4 on the Geometry subscale, 5 on the Algebra subscale, and 4 on the Data subscale. Both tests yielded raw scores based on the total number of items correct. The tests were evaluated along with other predictive measures, including *PSGs* completed by teachers and achievement proficiency scores from the previous year. Data were analyzed to yield predictive validity evidence of relations to other variables. The criterion variable was 2011 achievement proficiency scores yielded by large scale assessments.

Table 3. Percent of Items for Each Subskill Area by States and Screening Tests

<table>
<tr><th>Subskill</th><th>Arizona</th><th>Pennsylvania</th><th>South Carolina</th><th>MAAPS Tests</th></tr>
<tr><td colspan="5">Reading</td></tr>
<tr><td>Vocabulary</td><td>7% - 56%</td><td></td><td>16% - 80%</td><td>27%</td></tr>
<tr><td>Comprehension</td><td>9% - 58%</td><td>40% - 60%</td><td>0% - 60%</td><td>37%</td></tr>
<tr><td>Interpretation</td><td>33% - 81%</td><td>40% - 60%</td><td>24% - 96%</td><td>37%</td></tr>
<tr><td colspan="5">Mathematics</td></tr>
<tr><td>Numbers</td><td>18%</td><td>18% - 22%</td><td>17%-20%</td><td>38%</td></tr>
<tr><td>Geometry</td><td rowspan="2">24%</td><td>15% - 20%</td><td>13%-16%</td><td>15%</td></tr>
<tr><td>Measurement</td><td>12% - 15%</td><td>17%-21%</td><td>12%</td></tr>
<tr><td>Algebra</td><td>26%</td><td>25% - 30%</td><td>27%-30%</td><td>19%</td></tr>
<tr><td>Data Analysis</td><td>18%</td><td>15% - 20%</td><td>17%-21%</td><td>15%</td></tr>
</table>

Use of the Screening Tests

The study involved teachers (n = 41) of 8th grade students (n = 388) from multiple districts in Pennsylvania, Arizona, and South Carolina. State leaders recruited at least one school district each from the top third, middle third, and bottom third in achievement proficiency test results in their state. This sampling system resulted in a large, diverse sample featuring sub-samples of students that were representative of the states from which they were drawn.

Teacher participants attended trainings in their states to complete the PSGs and to learn to administer the computer-based tests. During a window from December through mid-January, students in each of the classrooms completed the computer-based screening tests. This period was selected to mimic the likely time at which decisions would be made regarding which students would be selected for each achievement proficiency test option. As part of large scale achievement proficiency testing, each student also completed standardized tests in reading and mathematics. These tests were administered in March (Pennsylvania), April (Arizona), and May

(South Carolina). The tests were subsequently scored by each state's vendor and the results were obtained by the MAAPS representative in each state.

Evidence of Success in Using the Screening Tests

The best predictors of students' 2011 Achievement Test scores across states and content areas were always the students' 2010 Achievement Test scores. Table 4 depicts correlations, which could theoretically be as high as 1.00, that reflect the strength of the relationship between scores from the various predictors and the 2011 tests. (Bivariate correlations indicate the strength of the relationship between scores, and partial correlations are corrected for other predictors, to indicate the *unique* contribution of each predictor). The MAAPS tests shared bivariate correlations in the Very Large range in three of six cases, indicating that they could be suitable substitutes if previous year's test scores were not available. In four of six cases, the MAAPS tests added substantial information (partial $\geq$ .20) to that which was already available from state tests, and in three of six cases the PSGs also added substantial information.

Table 4. Correlations between Predictor Scores and 2011 Achievement Proficiency Scores

	Arizona		Pennsylvania		South Carolina	
	Pearson	**Partial**	**Pearson**	**Partial**	**Pearson**	**Partial**
Reading						
2010 Reading Achievement	.82*	.47*	.78*	.64*	.72*	.56*
MAAPS Test	.76*	.39*	.52*	-.04	.55*	.29*
Reading PSG	.67*	.20*	.57*	-.06	.58*	.32*
Mathematics						
2010 Math Achievement	.73*	.41*	.87*	.56*	.66*	.57*
MAAPS Test	.75*	.44*	.76*	.46*	.32*	-.07
Math PSG	.55*	.21*	.79*	-.04	.36*	.17

Note. $p < .05$, one-tailed.

Lessons Learned with the Screening Test Component of the Project

The MAAPS Consortium successfully developed computer-based screening tests in reading and mathematics that could efficiently supplement the previous year's achievement tests. Teacher rating measures, such as the PSGs, were also found to provide additional information to meet the call for multiple valid measures to identify students who should be eligible for an AA-MAS. While the relative merit of the supplementary information from these two screening methods remains to be determined, current findings indicate that both types of data are worth collecting when predicting future achievement test scores or proficiency status.

State Stories & Implications for Assessment of Students with Disabilities

MAAPS in Arizona

The state of Arizona was a unique state in this project because it was exclusively comprised of general education classrooms that included two target students with disabilities. As such, it provided a case example of the full inclusion model. In addition, we compared the predictive validity of MyiLOGS and SEC using their class-based indices to predict average class achievement on the state achievement for the state of Arizona—the only state that provided class-specific achievement data for students in participating classrooms. The SEC AI was previously identified as an OTL proxy (e.g., Kurz et al., 2010; Porter, 2002). The AI quantifies alignment based on overlap between an enacted curriculum matrix (established teacher self-report) and a general curriculum matrix (established by content experts on the basis of state-specific standards) at the intersection of topic and cognitive demand. Low alignment can thus be function of misalignment among topics covered, cognitive demands emphasized, or both.

The results of the alignment analyses indicated that the AI averages ranged between .14 and .16 across states. The differences in AIs by class type were not statistically significant. With respect to convergent validity, none of the correlations between MyiLOGS OTL indices and the AI were statistically significant. Given the hypothesized relations between content and quality-related OTL indices and the AI in the range of .10 and .30, the analyses suffered from low power and were thus subject to Type II errors. In short, the present results could not be used to determine convergent validity between the MyiLOGS OTL indices and the AI.

For purposes of predictive validity, Arizona Department of Education personnel provided class averages of the 2010-2011 state test for each class logged by participating teachers. The unit of analysis was kept at the class level due to the SEC being a class-level alignment index. Given the small sample size (N = 16), these analyses also suffered from low power and were thus subject to Type II errors. Despite low power, the results indicated several statistically significant correlations with medium effect sizes above .50. For the Arizona subsample, the SEC AI was negatively correlated with class achievement with $r = -.52$ ($p < .05$). This finding is surprising given prior research findings, which have supported a positive relation between the AI and student achievement (e.g., Kurz et al., 2010; Smithson & Collares, 2007). An important difference between this subsample and samples in other predictive studies such as the ones in Kurz et al. (2010) is the sample's sensitization to their daily instructional practices. That is, teachers in this study reviewed their daily instruction several times a week for up to eight months prior to taking the SEC's annual survey. However, the extent to which this sensitization increased or decreased the accuracy with which teachers were completing the SEC's annual survey is unclear.

Three class-based OTL indices showed statistically significant relations with class achievement: *Time on Standards*, the *Cognitive Process Score*, and the *Grouping Format Score*. First, the average number of minutes per day dedicated to the state-specific standards had a positive relation with class achievement with a medium effect size. Second, a greater emphasis on high-order thinking skills correlated positively with class achievement also with a medium effect size. Third, a greater emphasis on small group and individual grouping formats correlated negatively with class achievement with a medium negative effect size. The latter finding is also surprising given prior research indicating a positive relation between achievement and grouping formats other than whole class (e.g., Elbaum et al., 2000). In addition, this finding cannot be attributed to class type—the prevalence of alternative grouping formats in special education classroom, which may further coincide with lower academic achievement—because the Arizona subsample was entirely comprised of general education classrooms.

In summary, the current analyses could not be used to substantiate convergent validity between the SEC AI and the MyiLOGS OTL indices. To do so, further research properly powered to detect the hypothesized relations is needed. With respect to the predictive validity of two class-based OTL indices—*Time on Standards* and the *Cognitive Process Score*—evidence was found to support their relation to class achievement.

Based on the available data, we further examined the relation between student-based OTL indices and individual student achievement for the Arizona subsample (N = 32). To this end, we applied several multiple regression models predicting current student achievement and three sets of time, content, and quality-related OTL indices. Without controlling for prior achievement, instructional time on custom skill/activities (*Time on Custom*) was the only student-based OTL index that exhibited a positive relation with student achievement accounting for about 24% of the variance. This finding is surprising in the context of a non-significant finding for *Time on Standards*. That is, one would expect that more instructional time on the state-specific standards would be related to higher achievement based on an assessment that covers those standards—rather than an index related to instructional time on objectives/activities outside the standards. However, as noted previously, many teachers logged review activities and technology-based elements of their lesson under *Time on Custom*. As such, it is very likely that *Time on Custom* reflected *additional* time on standards-based instruction rather than instructional time unrelated to the general curriculum standards.

None of the student-based OTL indices in the various models were significant predictors above and beyond students' prior achievement. An exploratory analysis using three models of student-based OTL quality indices for the various cognitive processes, instructional practices, and grouping formats indicated a statistically significant relation with student achievement for two instructional practices, *Elicited Think Aloud* and *Used Independent Practice*, as well as the *Whole Class* grouping format.

MAAPS in Pennsylvania

The MAAPS project brought together representatives of the Pennsylvania Department of Education, Bureau of Special Education; faculty from the Special Education Program in the School of Education at the University of Pittsburgh; and general and special education teachers and supervisors from numerous districts geographically spread throughout the Commonwealth. State personnel began recruiting school district and middle school teachers as soon as the school year was underway in the fall of 2010. To be included in the MAAPS project, each general and special education teacher had to provide Mathematics (MA) or English/Language Arts (ELA) instruction to at least two 8th-grade students with disabilities who had been or were likely to be assigned by their IEP team to participate in the newly launched Pennsylvania modified accountability assessment (the AA-MAS in Pennsylvania was designated as the Pennsylvania System of School Assessment [PSSA] -M). The PSSA-M had been introduced in PA in math only, for grades 4-8 and 11, in time for the spring testing in 2010. The state plan called for a build out of the AA-MAS to include reading and science as well as math for all tested grades (except third grade) in time for the annual spring assessment for the 2010-2011 school year. The math PSSA-M experience in the spring of 2010 had raised many questions about the manner in which IEP teams went about assigning students to the PSSA, the PSSA-M, and the alternate assessment, the PASA. Bureau of Special Education representatives and University of Pittsburgh faculty were eager to determine whether an additional screening measure might make the IEP team assignment task more objective and appropriate. Further, there remained a good deal of controversy in the Commonwealth regarding access to the general education standards and assessment anchors for students with disabilities, whether they were being taught in inclusive or pull-out service delivery models. Use of MyiLOGS promised to illuminate this issue by providing a comprehensive description of eighth-grade students' instructional access to the general curriculum standards and assessment anchors and the extent to which teachers were differentiating opportunity to learn for individual students with disabilities in their classrooms.

By the beginning of October, 5 middle schools and 12 teachers (7 general education teachers and 5 special education teachers) were committed to participate in the project. Trainings, fully implemented by mid-October, demonstrated that all 12 teachers could be trained to use MyiLOGS to criterion during the allotted 4.5-hour training time. Teachers rated the training, the instructional scenarios, as well as the allocated training time as helpful and sufficient for understanding how to use the MyiLOGS system. Once training was completed, the required logging period for all teacher participants was four full months with the option to continue through the month of April 2011. Target students were also required to take the online screening measures, the results of which might improve decision-making in assigning students to the general spring assessment or the modified assessment.

Although the teachers were a homogeneous group (11 females, 1 male; all Caucasian; 10 with Masters degrees, 2 with Bachelor) of experienced teachers (mean length of service of 10 years), the student body was diverse. Of the 19 students on whom MyiLOGS and screening data were collected, 10 were female, 12 were Caucasian, 14 were classified as having a learning disability, and 9 were eligible for free or reduced lunch.

Very early in the implementation of the MAAPS project, the Commonwealth, following guidance from the USDE, decided that despite the development and production work that had gone into the state's modified assessments (PSSA-Ms in reading, math, and science), spring 2012 would be the last year in which they would be administered. "Modified" accountability assessments were to be discontinued. As a result, attention to the screening measures aspect of MAAPS was minimal, and the full focus of the Pennsylvania group was on the feasibility and usefulness of the MyiLOGS tool.

The collected evidence substantiated that 8th grade general education and special education reading and math teachers could maintain high procedural fidelity logging various opportunity to learn indices at the class and student level across several months of a school year, and that teachers' log data provided a valid representation of their classroom instruction based on agreement percentages between teachers and independent observers.

Based on this sample's general education classrooms, which represented a full inclusion model, students with disabilities experienced less time on standards, more non-instructional time, and less content coverage compared to their classmates. These findings do not support a commonly held assumption in OTL research, namely that class-based OTL indices are sufficient for describing OTL of all students nested within that class. At least for students with disabilities nested in general education classrooms, OTL appears to be a differentiated opportunity structure. Secondly, the findings raise concerns that students with disabilities may not receive equal, let alone equitable, OTL compared to their class along several instructional dimensions. The instructional differences do not indicate equal or equitable OTL for students with disabilities.

But the real value of MyiLOGS for PA teachers was that it served as a self-monitoring tool for the teachers themselves. Daily or weekly recording of what was taught cross-referenced to the PA academic standards and anchors provided a kind of progress-monitoring feedback to the teachers on whether and how they were providing instructional opportunities that might impact students' performance on the annual accountability assessment. In combination with the documentation of instructional features such as grouping, cognitive challenge, and on-task behavior, the teacher logs provided a heretofore unavailable record of teacher self-report data on opportunity to learn indices at both the class and individual student levels.

As the potential value of MyiLOGS as a self-monitoring tool for teachers of students with mild/moderate disabilities who take the regular statewide assessment or the modified grade-level assessment became apparent, the PA group decided to explore whether the tool might also have value in helping teachers of students with the most significant cognitive disabilities monitor the opportunities they were providing their students to learn the alternate assessment anchors and standards that undergird the PA alternate assessment. For this project, we focused on reading and used a set of reading skills not the grade-level alternate standards. We decided that more discreet skills (e.g., phonological awareness, sight word instruction, and fluency) would better allow us to describe instruction. The alternate standards were considered too broad.

We recruited 19 teachers of 7th and 8th grade students who take the Pennsylvania Alternate System of Assessment to use MyiLOGS in conjunction with early reading CBM measures for about 18 weeks. Each teacher administered the reading measures to at least two students, used an online opportunity to learn (OTL) tool loaded onto an iPad to document aspects of their reading instruction, and provided feedback on their use of the assessment data and the OTL tool. Students completed weekly CBM reading measures (two 1-minute timed assessments of word and passage reading).

The teachers in this build-out of the MAAPS project were responsible for implementing evidence-based reading instruction for students with significant intellectual disabilities (ID). Despite the high priority PA places on inclusive service delivery models for all students with disabilities, these teachers taught in traditional self-contained classrooms. This project examined the types and frequency of reading instruction provided to this population of students through documentation on a "cool" piece of handheld technology. The purpose of this study was to document the reading instruction being provided to students with significant cognitive disabilities and to determine if MyiLOGS and CBM were perceived by teachers as useful for enhancing their planning for and implementing of reading instruction. Though the study has not been completed as of this date (June, 2012), feedback from the teachers confirms the usefulness of the MyiLOGS tool as an efficient and often vivid feedback mechanism with which teachers can monitor the progress of their reading instruction and the opportunities to learn they are proving their students.

Based on the results of the MAAPS project in Pennsylvania, we learned a number of lessons that will have implications for advancing assessment and instructional practices for students with disabilities. The lessons are:

1. Despite detailed guidance and explicit enrollment criteria provided by the Pennsylvania Department of Education, Bureau of Special Education, IEP teams made very idiosyncratic decisions about who to assign to take the PSSA-M. This is a moot point, however, since

the modified assessment is being phased out. Decisions of IEP teams regarding who should be assigned to take the alternate assessment based on alternate achievement standards are both more clear-cut and more consistent.

2. While the federal guidance regarding assignment of students to a modified assessment suggests the review of multiple sources of achievement data, students with disabilities who are accessing the general education curriculum and will take the regular or modified accountability assessment perform academic skills with a high degree of consistency. Therefore, additional test evidence, beyond last years' performance on the statewide assessment, is redundant.

3. MyiLOGS provides valid documentation of the instructional opportunities being provided to students with disabilities by their general or special education teacher. It is useful in documenting instructional opportunities for students with mild/moderate disabilities and can also be useful for teachers of students with significant cognitive disabilities in the early stages of reading instruction.

4. Perhaps one of the most interesting uses of MyiLOGS is the explicit and concrete feedback the data provide to teachers about what they are doing in the course of instruction. Teachers could use this feedback in the same way that they use progress monitoring tools for students, to monitor their own instruction and make deliberate and planned changes in instruction on the basis of the MyiLOGS data.

5. With shrinking school budgets, teachers are being asked to do more with less in their day-to-day teaching. Time unfortunately has become a commodity that today's classroom teachers are lacking. In addition to daily lessons, special education teachers must manage enormous amounts of paperwork and document progress on individualized education programs (IEP) for every student in their class. Educational tools are needed that will assist special education teachers in maximizing their limited amount of instructional time. MyiLOGS may be one solution, and MyiLOGS on an iPad (instead of a desktop or laptop computer) make the tool even easier to use. Not only are tablets portable and easy to manipulate, they provide users an array of functions that can be tailored to their individual needs. The MAAPS project underscored the need for additional research that examines how handheld technology may assist special education teachers to be more efficient and better teachers.

MAAPS in South Carolina

The teacher composition in South Carolina for the MAAPS OTL Study was similar to Pennsylvania with a mixture of general and special education classes. As such, the data from South Carolina resembled the data from Pennsylvania. However, the effect sizes for Instructional Time on Standards, Non-Instructional Time, and Content Coverage were slightly larger.

Teachers in South Carolina were particularly supportive of the usability and promise of the tool and provided valuable feedback during the development phase of MyiLOGS through the MAAPS Pilot Study. They further asked for additional functionalities to increase the tool's utility such as lesson planning tool.

Lastly, South Carolina extended the use of MyiLOGS through a supplemental study during the no-cost extension phase into an itinerant teacher population for students of visual impairments. The application of MyiLOGS was supported by the *The Vision Institute* (TBI) of South Carolina and featured a specialized intended curriculum called the Extended Core Curriculum. The online teacher logging capability allowed the collection of data from this remotely located and dispersed teacher population. In addition, the calendar interface was modified to feature daily instructional details. This teacher population confirmed the system's usability for daily instructional logging.

Conclusions from MAAPS and Next Steps

The MAAPS Project was a successful multi-state project that advanced work with 8th grade students with disabilities who had a history of poor performances on their statewide achievement tests. The main goal of the project was to develop a system—the *Modified Alternate Assessment Participant Screening system*—that could help educators on IEP teams accurately identify students likely to qualify for an AA-MAS. The parts of the system were (1) academic screening tests in math and language arts aligned with state content standards, and (2) a measure of opportunity to learn MyiLOGS. We found that both of these tools for measuring learning opportunities and academic achievement could be used to predict end of year performance on state tests reasonably well, but not substantially better than using only students' previous years performances on the state test. However, in the spirit of innovation, multiple measurement, and convergent evidence for decision making, the MAAPS screening test and MyiLOGS proved promising tools or methods for educators.

Findings from MyiLOGS portion of this project, in particular, have alerted many educators and fellow researchers to seriously focus more on addressing issues of instruction and students' opportunity to learn the content emphasized in the intended curriculum and tested on state achievement tests. As we learned, one of the pervasive reasons that a substantial portion—well more than 2%—of students with disabilities do not perform proficiently on state achievement tests is likely the result of not having the opportunity to learn the content in their classrooms during the 140+ days leading up to the test. The findings from the MAAPS project adds to our belief that more must be done instructionally to advance the learning of the knowledge and skills we value for all students. Students with disabilities often learn at a slower pace than their peers without disabilities, so logically they will need more instructional time to cover the content of the intended curriculum. If this does not happen, their potential will remain invalidly measured and under-represented.

Resources

Readings & A MyiLOGS User Guide

Downing, S. (2006). Twelve Steps for Effective Test Development. In S. Downing & T. Haladyna (Eds.), *Handbook of Test Development*. Mahwah, NJ: Lawrence Erlbaum Associates.

Elliott, S. N., & Kurz, A. *MyiLOGS Users Guide.* Available to download from http://lsi.asu.edu/project/modified-alternate-assessment-participation-screening-maaps-consortium

Feeney-Kettler, K. A., Kratochwill, T. R., & Kettler, R. J. (2011). Identification of preschool children at risk for emotional and behavioral disorders: Development and validation of a universal screening system. *Journal of School Psychology, 49*(2), 197-216.

Lemons, C. J., Kloo, A., & Zigmond, N. (2011). Implementing modified achievement tests: Questions, challenges, pretending, and potential negative consequences. In S. N. Elliott, R. J. Kettler, P. A. Beddow & A. Kurz (Eds.), *Handbook of accessible achievement tests for all students: Bridging the gaps between research, practice, and policy* (pp. 295-317). NY: Springer.

Zigmond, N., Kloo, A., & Lemons, C. J. (2011). IEP Team Decision-Making for More Inclusive Assessments: Policies, percentages, and personal decisions. In S. Elliott, R. Kettler. P. Beddow, & A. Kirz (Eds), *Handbook of Accessible Achievement Tests for All Students* (pp. 69-82). New York: Springer

Zigmond, N., & Kloo, A. (2009). The “two percent students”: Considerations and consequences of eligibility decisions. *Peabody Journal of Education, 84*(4), 478-495.

References

Abedi, J., Courtney, M., Leon, S., Kao, J. C., & Azzam, T. (2006). *English language learners and math achievement: A study of opportunity to learn and language accommodation* (CRESST Report No. 702). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Aguirre-Munoz, Z., Boscardin, C. K., Jones, B., Park, J., Chinen, M., Shin, H. S., Benner, A. (2006). *Consequences and validity of performance assessment for English language learners: Integrating academic language and ELL instructional needs into opportunity to learn measures* (CSE Report No. 678). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Anderson, L. W. (1986). Opportunity to learn. In T. Husén & T. Postlethwaite (Eds.), *International encyclopedia of education: Research and studies* (pp. 3682-3686). Oxford, UK: Pergamon.

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Rath, J., & Wittrock, M. C. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.

Bloom, B. S. (1976). *Human characteristics and school learning*. New York: McGraw-Hill.

Borg, W. R. (1980). Time and school learning. In C. Denham & A. Lieberman (Eds.), *Time to learn* (pp. 33-72). Washington, DC: National Institute of Education.

Brophy, J., & Good, T. L. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), *Handbook of Research on Teaching* (3rd ed., pp. 328-375). New York: Macmillian.

Elbaum, B., Vaughn, S., Hughes, M., T. Moody, S. W., & Schumm, J. S. (2000). How reading outcomes for students with learning disabilities are related to instructional grouping formats: A meta-analytic review. In R. Gersten, E. P. Schiller, & S. Vaughn (Eds.), *Contemporary special education research: Syntheses of the knowledge base on critical instructional issues* (pp. 105-135). Mahwah, NJ: Lawrence Erlbaum.

Elliott, S. N., & Gresham, F. M. (2007). *SSiS: Classwide Intervention Program.* Bloomington, MN: Pearson Assessments.

Foorman, B. R., Francis, D. J., Winikates, D., Mehta, P., Schatschneider, C, & Fletcher, J. M. (1997). Early intervention for children with reading disabilities. *Scientific Studies of Reading, 1*, 255-276.

Gersten, R., Chard, D. J., Jayanthi, M., Baker, S. K., Morphy, P., & Flojo, J. (2009). Mathematics instruction for students with learning disabilities: A meta-analysis of instructional components. *Review of Educational Research, 79*(3), 1202-1242.

Gettinger, M., & Seibert, J. K. (2002). Best practices in increasing academic learning time. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV* (Vol. 1, pp. 773-787). Bethesda, MD: National Association of School Psychologists.

Herman, J. L., & Abedi, J. (2004). *Issues in assessing English language learners' opportunity to learn mathematics* (CSE Report No. 633). Los Angeles: Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing.

Karvonen, M., Wakeman, S. Y., Flowers, C., & Browder, D. M. (2007). Measuring the enacted curriculum for students with significant cognitive disabilities: A preliminary investigation. *Assessment for Effective Intervention*, *33*, 29-38.

Kettler, R. J. (2011). Computer-based screening for the new modified alternate assessment. *Journal of Psychoeducational Assessment*, *29*(1), 3-13.

Kettler, R. J., Elliott, S. N., Davies, M., & Griffin, P. (2012). Using academic enabler nominations and social behavior ratings to predict students' performance on Australia's national achievement test. *School Psychology International*, *33*, 93-111.

Klingner, J. K., Vaughn, S., Hughes, M. T, Schumm, J. S., & Elbaum, B. (1998). Academic outcomes for students with and without learning disabilities in inclusive classrooms. *Learning Disabilities Research and Practice*, *75*(3), 153-161.

Kurz, A. (2011). Access to what should be taught and will be tested: Students' opportunity to learn the intended curriculum. In S. N. Elliott, R. J. Kettler, P. A. Beddow, & A. Kurz (Eds.), *Handbook of accessible achievement tests for all students: Bridging the gaps between research, practice, and policy* (pp. 99-129). New York: Springer.

Kurz, A., & Elliott, S. N. (2011). Overcoming barriers to access for students with disabilities: Testing accommodations and beyond. In M. Russell (Ed.), *Assessing students in the margins: Challenges, strategies, and techniques* (pp. 31-58). Charlotte, NC: Information Age Publishing.

Kurz, A., Elliott, S. N., & Kettler, R. J. (2012). A*ssessing students' opportunity to learn the intended curriculum: Initial validity evidence for an online teacher log*. Manuscript submitted for publication.

Kurz, A., Elliott, S. N., Wehby, J. H., & Smithson, J. L. (2010). Alignment of the intended, planned, and enacted curriculum in general and special education and its relation to student achievement. *Journal of Special Education*, *44*, 131-145.

Lyon, G. R., Fletcher, J. M., Fuchs, L. S., & Chhabra, V. (2006). Learning disabilities. In E. J. Mash (Ed.), *Treatment of childhood disorders* (Vol. 3, pp. 512–591). New York: Guilford Press.

Marzano, R. J. (2000). *A new era of school reform: Going where the research takes us* (REL No. #RJ96006101). Aurora, CO: Mid-continent Research for Education and Learning.

Mayer, R. E. (2008). *Learning and instruction* (2nd ed.). Upper Saddle River, NJ: Pearson.

Porter, A. C. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher*, *31*(7), 3-14.

Smithson, J. L., & Collares, A. C. (2007, April). *Alignment as a predictor of student achievement gains*. Paper presented at the meeting of the American Educational Research Association, Chicago, IL.

Stevens, F. I., & Grymes, J. (1993). *Opportunity to learn: Issues of equity for poor and minority students* (NCES No. 93-232). Washington, DC: National Center for Education Statistics.

Swanson, H. L. (2000). What instruction works for students with learning disabilities? Summarizing the results from a meta-analysis of intervention studies. In R. Gersten, E. P. Schiller, & S. Vaughn (Eds.), *Contemporary special education research: Syntheses of the knowledge base on critical instructional issues* (pp. 1-30). Mahwah, NJ: Lawrence Erlbaum.

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Lindamood, P., Rose, E., Conway, T., & Garvan C. (1999). Preventing reading failure in young children with phonological processing disabilities: Group and individual differences to instruction. *Journal of Educational Psychology, 91*, 579–593.

U.S. Department of Education. (April, 2007). *Modified Academic Achievement Standards: Non-Regulatory Guidance.* Washington, D.C.: Author.

Vaughn, S., Gersten, R., & Chard, D. J. (2000). The underlying message in LD intervention research: Findings from research syntheses. *Exceptional Children, 67*(1), 99-114.

Walberg, H. J. (1986). Syntheses of research on teaching. In M. C. Wittrock (Ed.), *Handbook of Research on Teaching* (3rd ed., pp. 214-229). New York: Macmillian Publishing Company.

Wang, J. (1998). Opportunity to learn: The impacts and policy implications. *Educational Evaluation and Policy Analysis, 20*(3), 137-156.

Webb, N. L. (2006). Identifying content for student achievement tests. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of Test Development* (pp. 155-180). Mahwah, NJ: Lawrence Erlbaum.