EA21-638-2

2020-21

Evaluation of Assessment for Learning Program

DEPARTMENT OF EVALUATION AND ASSESSMENT

Michael Hinojosa, Ed. D.
Superintendent of Schools

DALLAS INDEPENDENT SCHOOL DISTRICT

2020-21

Evaluation of Assessment for Learning Program

Approved Report of the Department of Evaluation and Assessment

EA21-638-2
August 2021

C. Elizabeth Watson, M.P.A., Evaluation Analyst II

Anduamlak Meharie, Ph.D.
Manager, Program Evaluation

Joan Bush, Ph.D.
Executive Director, Program Evaluation

Cecilia Oakeley, Ph.D.
Assistant Superintendent, Evaluation and Assessment

© 2021 Dallas Independent School District (Dallas ISD)
Mail a written request for permission to use or reproduce any part of this document to:
Dallas ISD Department of Evaluation and Assessment
H. B. Bell School Support Service Center; 2909 N. Buckner Blvd, Box 10; Dallas, TX 75228.

This report and executive summary are available for download at www.dallasisd.org/Page/888. For more information, please contact Program Evaluation at evaluation@dallasisd.org.

TABLE OF CONTENTS

EXECUTIVE SUMMARY 1
PROGRAM DESCRIPTION 5
PURPOSE AND SCOPE OF THE EVALUATION 5
EVALUATION QUESTIONS AND RESULTS 6
What are the benefits of formative assessment for instructional practices? 6
What was the amount of funding for the program? 7
Methodology 7
Results 7
How was the Assessment for Learning program implemented? 7
Methodology 7
Results 7
What were teacher perceptions of the Assessment for Learning program? 10
Methodology 10
Results 11
How did participation in the Assessment for Learning program affect student achievement? 12
Methodology 12
Results 13
What are the plans for the future of the Assessment for Learning program? 17
Methodology 17
Results 17
KEY FINDINGS AND RECOMMENDATIONS 18
Key Findings 18
Recommendations 19
REFERENCES 20
APPENDICES 21

LIST OF TABLES

Table 1: 2020-21 *MAP* English Reading Assessment Mean RIT Scores by Grade, Students of Teachers Implementing Assessment for Learning 14
Table 2: 2020-21 *MAP* English Reading Assessment Mean RIT Scores by Grade, Comparison Group of Students of Teachers Not Implementing Assessment for Learning 14
Table 3: 2020-21 *MAP* Mathematics Assessment Mean RIT Scores by Grade, Students of Teachers Implementing Assessment for Learning 15
Table 4: 2020-21 *MAP* Mathematics Assessment Mean RIT Scores by Grade, Students of Teachers Not Implementing Assessment for Learning 15
Table 5: 2020-21 Dallas ISD *MAP* Reading (English Only) and Mathematics Growth Threshold Rates by Assessment for Learning (AfL) Comparison Group: BOY to MOY 16
Table 6: 2020-21 Dallas ISD *MAP* Reading (English Only) and Mathematics Growth Threshold Rates by Assessment for Learning (AfL) Comparison Group: MOY to EOY 17

FIGURE

Figure 1: 2021 Assessment for Learning Teacher Survey Responses 11

LIST OF APPENDICES

A FARROP Rubric 22
B Assessment for Learning Teacher Survey Results 32

ACRONYMS AND ABBREVIATIONS

AfL	Assessment for Learning
BOY	Beginning-of-year
CGP	Conditional Growth Percentile
EOY	End-of-year
FARROP	Formative Assessment Rating Rubrics and Observation Protocol
ISD	Independent School District
MAP	*Measures of Academic Progress*
MOY	Middle-of-year
PLC	Professional Learning Community
POD	Plan, Observe, Debrief
RIT	Rasch Unit

2020-21

Evaluation of Assessment for Learning Program

DEPARTMENT OF EVALUATION AND ASSESSMENT

EXECUTIVE SUMMARY

The purpose of this report is to provide context and process evaluation information related to key components of the Assessment for Learning program. The focus of the 2020-21 evaluation was to provide context and determine the level of implementation of the grant components, perceptions of staff who participated in grant activities, and the impact of the program on student achievement. Data for this report were obtained from interviews with program staff members, a review of internal documents, and analyses of available survey and student achievement data. Dallas ISD used $186,357 in Title II grant funds to implement the program in 2020-21. The program hired a coordinator to directly support campuses, lead professional development sessions, and coach teachers and campus leaders.

The focus of the 2020-21 evaluation was to determine the level of implementation of the grant-funded programs and to examine participant perceptions of the programs, as applicable. Data for this report were obtained from interviews with program staff members assigned to each grant component, a review of internal documents, and analyses of available survey data.

EVALUATION QUESTIONS/OUTCOMES

How was the Assessment for Learning program implemented?

Program Characteristics

In 2021-21, the Assessment for Learning program took place on eleven campuses: Ben Milam, Kiest, Zaragoza, Stevens Park, Connor, and Tatum Elementary Schools; Franklin and Spence Middle Schools; and Woodrow Wilson, Molina, and Adamson High Schools. A total of 139 teachers, across all grade levels and content areas, implemented the program in their classrooms. Assessment for Learning utilized an innovative and highly communicative approach to teaching that emphasizes formative assessment throughout instruction.

Formative assessment is a planned, ongoing process used by students and teachers during learning and teaching to elicit and use evidence of student learning to improve student understanding of intended content area learning outcomes and support students to become self-directed learners. Effective use of the formative assessment process requires students and teachers to integrate and embed the following practices in a collaborative and respectful classroom environment: clarifying learning goals and success criteria within a broader progression of learning; eliciting and analyzing evidence of student thinking; engaging in self-assessment and peer feedback; providing actionable feedback; and using evidence and feedback to move learning forward by adjusting learning strategies, goals, or next instructional steps (Fisher, Frey, & Hattie, 2018). Teachers operate in daily feedback loops where students and teachers answer the following questions: Where am I going? Where am I now? How do I close the gap? or Where to next? (Sadler, 1998).

Campus Walks and Observations

Assessment for Learning program staff utilize the Formative Assessment Rating Rubrics and Observation Protocol (FARROP) to determine implementation of the ten dimensions of formative assessment in classrooms. The Assessment for Learning coordinator conducted a total of four campus walks throughout the year (November, February, March, and May). Because the program's results were noted informally and did not involve collecting quantitative measures, an analysis of results was not possible.

In addition, program staff and campus leaders provided coaching and training for teachers in the year three pilot program through observations and POD (Plan, Observe, Debrief) cycles. A POD cycle is a one-and-a-half-day process where teachers plan together, observe each other's classrooms, and receive feedback from one another and their trainer. In previous years, the program conducted formal POD cycles in November, January, and March. However, due to the shortage of substitute teachers in 2020-21, formal POD cycles were impossible. Instead, campuses engaged in "mini-PODs" where they used conference or PLC time to observe each other. At other campuses, teachers recorded video of lessons and used the POD protocol to give each other feedback after viewing.

Teacher Training

The Assessment for Learning program provided several traditional professional development opportunities throughout the year in addition to the ongoing observation and coaching cycles. A total of 1,137 district teachers, principals, and central staff members attended an online presentation by Doug Fisher on September 1, 2020. A total

of 2,532 district teachers, principals, and central staff members attended a professional development session led by formative assessment expert John Hattie on January 4, 2021. Further, 4,236 district teachers, principals, and central staff members attended training with John Almarode on improving distance learning, on March 10, 2021. Assessment for Learning teachers also had access to asynchronous digital training materials throughout the year via Schoology, a learning management system managed by PowerSchool.

Simultaneous Learning

Teachers within the Assessment for Learning program implement simultaneous learning, with distance and in-person students in the same classroom at the same time. Through this approach, the program seeks to create an equitable environment through shared culture, access, and expectations for both in-person and distance learners. Teachers were trained to design the physical classroom environment to support both distance and in-person learners while selecting appropriate videoconferencing tools to leverage screen sharing for equitable access. Teachers established norms and expectations for both cohorts of students without attempting to plan two different courses.

Simultaneous Learning emphasized providing clear learning goals, success criteria, and expectations and building in opportunities for collaboration, feedback, and assessment. Teachers explained the learning goals, success criteria, norms, and expectations. They emphasized assigning student roles to promote agency and collaboration, while building in opportunities for teacher- and student-led small groups. Teachers were encouraged to intentionally create moments of relationship-building within and across modalities of in-person and distance learning.

What were teacher perceptions of the Assessment for Learning program?

Methodology

The Assessment for Learning program utilized Google Forms to craft and administer a feedback survey exploring teachers' evolving implementation of Assessment for Learning practices. The program manager provided an online survey link to Assessment for Learning teachers at three different points throughout the year. The beginning-of-year (BOY) survey link was open from September 10, 2020 to November 17, 2020. The middle-of-year (MOY) survey link was open from January 23, 2021 to March 7, 2021. The end-of-year (EOY) survey link was open from May 24, 2021 until June 18, 2021. The purpose of the survey was to collect feedback related to participants' individual experiences working in the Assessment for Learning program. The survey included one demographic item and 10 items with Likert-type response options rating staff perceptions of the program. All Assessment for Learning teachers (n = 124) were selected to receive the survey. Program staff shared raw survey data with the program evaluator. The evaluator used frequency analyses to determine rates of positive responses for each scaled item, along with the number and percentage of responses per category for other items.

Results

A total of 115 respondents completed at least one survey, resulting in a 93 percent response rate, though participation dropped sharply as the year progressed (61 teachers responded during the MOY survey window, and 12 teachers responded during the EOY survey window.)

Overall, the proportion of teachers who indicated that they incorporated each behavior into their teaching increased over the course of the school year. The largest increases were seen in posting success criteria that tells students what successful learning will look like (61.8% at BOY versus 83.3% at EOY, an increase of 21.5 percentage points) and providing opportunities for students to set goals, reflect, and monitor their own progress (42.6% at BOY versus 75.0% at EOY, an increase of 32.4 percentage points). At the beginning of the year, 76.5 percent of teachers reported that they modeled how to use mistakes, failure, and struggle as opportunities for growth. By the end of the year, 91.7 percent of teachers reported "often" or "usually" doing so. The formative assessment skill rated least implemented at the beginning of the year (28.7% at BOY), teaching students how to give and receive feedback, increased by 13 percentage points by the end of the year (41.7% at EOY).

How did participation in the Assessment for Learning program affect student achievement?

Methodology

The NWEA *Measures of Academic Progress* (*MAP)* assessments are universal screeners administered to Dallas ISD students in grades one through ten at the beginning (BOY), middle (MOY), and end (EOY) of the school year. *MAP* data consists of several variables used to describe student performance and academic growth over time by subject. The Rasch Unit (RIT) score represents the level at which students are ready to learn at the time of assessment. The RIT score is a standardized, equal-interval score that can be compared to other RIT scores regardless of student age or grade. The second MAP Growth variable used for this report was the Conditional Growth Percentile (CGP). The CGP represents the extent to which student achievement changed from one testing window to the next, relative to change for other students who are similar in starting achievement level,

grade, and number of instructional weeks between testing windows. For each subject assessed, CGPs for each student are calculated with respect to a large, nationally-normed sample obtained by NWEA and can be calculated whenever a student takes a *MAP* assessment in a new testing window (e.g., BOY to MOY, MOY to EOY, BOY to EOY). A CGP at the 50th percentile indicates that the RIT score increase from the prior testing window represents the median amount, or expected level, of progress for that grade level[1].

The evaluator used student-level *MAP* data files dated November 5, 2020 (BOY), March 30, 2021 (MOY), and June 23, 2021 (EOY) and merged them with the district demographic data files pulled closest to the dates of the *MAP Growth* data files for each testing window. The evaluator selected a sample of 50 grades one through 11 Assessment for Learning teachers. Students were flagged as participants if they were identified by district course files as being taught by one of these teachers. To construct comparison groups of comparable classrooms not implementing the program, classrooms of the same grade on the same campus as implementing teachers were chosen whenever possible. In the case of Ben Milam Elementary, where the entire campus was implementing the program, eight teachers from the same grade levels at Lipscomb Elementary were selected. This campus was chosen based on similarity of campus-level demographic and achievement criteria used to create state accountability comparison groups. Students were selected into the sample if they had valid *MAP* scores for all three test administrations (English language only in Reading, as English and Spanish MAP Reading assessments are not directly comparable). In the cases where students appeared on rosters for both AfL and non-AfL teachers (such as high school students taking courses from multiple teachers on a campus), the student was omitted from the study. The evaluator then used frequency and crosstabs analyses to calculate the average *MAP* RIT and CGP scores by subject, grade, and student group. The evaluator conducted *t*-tests of mean RIT scores to ascertain if the program made a statistically significant difference in the participants' academic growth as measured by MAP results. Also, the evaluator used Cohen's *d* to examine if differences were practically meaningful. Finally, the evaluator conducted chi-square tests of independence to determine if a significant association existed between AfL participation and growth threshold rates as illustrated by the CGP score.

Results

At all grade levels, RIT scores for reading and mathematics generally remained unchanged or increased from BOY to MOY and from MOY to EOY. For each assessment and testing period, mean RIT scores generally increased for higher grade levels, as expected due to the nature of the assessment (scores are given on a single continuum of RIT scores from grades one through 11).

While AfL students' overall mean Reading RIT across all grades was five points higher than non AfL students at BOY (210 versus 205), the mean RIT for Afl students increased by three points by MOY (210 to 213), while increasing by four points for non-AfL students (205 to 209). From MOY to EOY, AfL students across all grades lost the previous gains, with an overall mean RIT score of 210, the same as at BOY. For non-AfL students, the overall mean EOY RIT score was three points higher than at BOY (205 versus 208). Notably, students at grades one through three saw larger gains from BOY to EOY than upper grades: 11 to 18 points for AfL students and nine to 15 points for non-AfL students. In addition, AfL students at grade eight saw a loss of six points (206 to 200) from BOY to EOY, while non-AfL students gained one point during this time period (227 to 228).

While there were slight differences in mean RIT scores, *t*-test and effect size analyses of the differences between AfL and non-AfL students' mean RIT Reading scores at BOY, MOY, and EOY showed no statistical or practical significance.[2] For BOY, t(20) = 0.17, $p = 0.86$, ***Cohen's d*** = 0.19 . For MOY, t(20) = 0.22, $p = 0.82$, ***Cohen's d*** = 0.18. For EOY, t(20) = 0.28, $p = 0.78$, ***Cohen's d*** = 0.10.

Overall mean Mathematics RIT scores for AfL students was three points higher at BOY than for non-AfL students (212 versus 209). From BOY to MOY, the mean RIT increased by eight points for both AfL students (212 to 220) and non-AfL students (209 to 217). From MOY to EOY, AfL students across all grades gained two points (220 to 222). For non-AfL students, the mean RIT score also increased by three points from MOY to EOY (217 to 220). Again notably, students at lower grades, this time one through three, saw notable gains from BOY to EOY, 15 to 23 points for AfL students and 11 to 16 points for non-AfL students. Again, *t*-test and effect size analyses of the differences between AfL and non-AfL students' mean RIT Mathematics scores at BOY, MOY, and EOY showed no statistical or practical significance. For BOY, t(20) = 0.37, $p = 0.72$, ***Cohen's d*** = 0.13. For MOY, t(20) = 0.28, $p = 0.78$, ***Cohen's d*** = 0.09. For EOY, t(20) = 0.32, $p = 0.76$, ***Cohen's d*** = 0.08.

To measure growth across testing periods, the evaluator divided CGP scores into three categories based on NWEA's definition of meeting expected growth. Students with a CGP below the 45th percentile "failed to meet" nationally-normed growth expectations. Those with a CGP between the 46th

[1] F7or more information, see https://community.nwea.org/docs/DOC-1630.
[2] A *p*-value less than 0.05 (typically ≤ 0.05) is statistically significant. A *p*-value higher than 0.05 (> 0.05) is not statistically significant. Cohen's *d* values of .20, .50, and .80 are considered small, medium, and large, respectively (Cohen, 1992).

and 55th percentile "met" expectations, and those with a CGP above the 55th percentile "exceeded" those expectations for growth.

For reading, the percentage of AfL students who met or exceeded growth thresholds by MOY was 55.4 percent which was lower by 2.5 percentage points compared to Non-AfL students (57.9%). For mathematics, the percentage of AfL students who met or exceeded growth thresholds by MOY was 71.1 percent which was higher by 1.7 percentage points compared to Non-AfL students (69.4%). There were no significant differences between students of AfL and Non-Afl teachers in meeting or exceeding growth expectations (2.5 percentage points or less). Further, a chi-square test of independence showed that there was no significant association between AfL participation and growth threshold rates in either Reading ($X^2 = 5.03$, $p = .08$) or Mathematics ($X^2 = 0.49$, $p = .78$).

For reading, the percentage of AfL students who met or exceeded growth thresholds by EOY was 34.2 percent which was lower by 4.8 percentage points compared to Non-AfL students (39.0%). For mathematics, the percentage of AfL students who met or exceeded growth thresholds by EOY was 49.4 percent which was lower by 1.5 percentage points compared to Non-AfL students (50.9%). There were no significant differences between students of AfL and Non-Afl teachers in meeting or exceeding growth expectations (4.8 percentage points or less). Further, a chi-square test of independence showed that there was no significant association between AfL participation and growth threshold rates in either Reading ($X^2 = 5.03$, $p = .08$) or Mathematics ($X^2 = 0.26$, $p = .88$).

What are the plans for the future of the Assessment for Learning program?

Methodology

Evaluation of the Assessment for Learning program included interviews with program management and a review of internal documents. The purpose of the interviews was to gather information about program requirements and activities. Follow-up information was collected by email.

Results

The Assessment for Learning Design Team was an advisory group consisting of program management, representatives from the Teaching and Learning and Evaluation and Assessment departments, AfL campus principals, and AfL teachers. The Design Team met four times during the year, in November, December, February, and June, providing feedback on program implementation, sharing campus and classroom challenges and solutions, and advising program management on program expansion plans.

On Saturday June 12, 2021, the program held an Assessment for Learning Summit on Zoom. The event lasted from 9:00 am to 1:00 pm and was open to all AfL teachers and principals. Three guest presenters joined district staff for author- and teacher-led breakout rooms, and discussions around books supplied to attendees upon registration. Formative assessment expert John Almarode addressed learning goals, success criteria, and feedback. Author Anthony Muhammad discussed equity and transforming school culture. Author and educational consultant Dr. Nicole Law presented on culturally responsive pedagogy and visible learning. A total of 198 district staff members attended.

While the program was implemented by at least some teachers at 11 campuses in 2020-21, the district is seeking to expand the program to include additional high priority campuses in 2021-22. Additionally, the program will expand to campus-wide implementation at several existing AfL campuses including Adamson and Wilson High Schools, Spence Middle School, and Conner, Stevens Park, Tatum, and Zaragoza Elementary Schools. Program management and district leadership are closely monitoring data to see if these practices could potentially be scaled district-wide as early as 2022-23.

RECOMMENDATIONS

Based on the key findings, the following recommendations are offered for stakeholders.

Formalize campus walk feedback.

Continue strong collaboration practices, including design team, PODs, and campus walks.

2020-21

Evaluation of Assessment for Learning Program

DEPARTMENT OF EVALUATION AND ASSESSMENT

PROGRAM DESCRIPTION

Dallas Independent School District (Dallas ISD) has worked closely with the Michael and Susan Dell Foundation, WestEd, and Education First to shift the thinking and practices around formative assessment in Dallas ISD through a pilot of the Assessment for Learning program with four schools beginning in 2018-19. These schools, Anne Frank, Ben Milam, Molina, and Benjamin Franklin included teachers in the program that spanned all grade levels and content areas intentionally incorporating formative assessment into their daily instruction with the support from coaching and professional development provided by program staff and experts in the field. During the third year of the program in 2020-21, the district scaled up the program to include teachers at seven additional campuses in the professional development and coaching process and continued to train leaders to better utilize formative assessment.

On March 13, 2020, in alignment with federal recommendations, the Governor of Texas issued a disaster proclamation certifying that the novel coronavirus COVID-19 posed an imminent threat of disaster for all counties in the state of Texas. In response to the COVID-19 pandemic, Dallas ISD transitioned students to at-home learning on March 23 following spring break (March 16 – 20) until the end of the 2019-20 school year. To help district students and families access the internet during at-home learning, the Dallas ISD launched "Operation Connectivity" in May 2020, which provided 23,000 hotspots to families by October 2020. In 2020-21, Dallas ISD delayed the start of the 2020-21 school year until September 8, 2020 and implemented at-home learning districtwide for the first four weeks. The district also modified the school calendar from a six-week schedule to a nine-week schedule. On October 22, 2020, the district announced that campuses would provide in-person and virtual learning throughout the rest of the school year, and parents could then choose whether their children participated in in-person or virtual instruction. In-person and virtual instruction included both synchronous and asynchronous learning activities. The district also continued to implement the Center for Disease Control (CDC) as well as state and local COVID safety protocols.

PURPOSE AND SCOPE OF THE EVALUATION

The purpose of this report is to provide context and process evaluation information related to key components of the Assessment for Learning program. The focus of the 2020-21 evaluation was to provide context and determine the level of implementation of the grant components, perceptions of staff who participated in grant activities, and the impact of the program on student achievement. Data for this report

were obtained from interviews with program staff members, a review of internal documents, and analyses of available survey and student achievement data.

EVALUATION QUESTIONS AND RESULTS

What are the benefits of formative assessment for instructional practices?

The research literature offers multiple, sometimes conflicting, definitions of formative assessment that evolved from a range of perspectives among teachers and school leadership. For example, formative assessment has been referenced as a process for making instructional adjustments based on feedback about student performance (Council of Chief State School Officers, 2007; Popham, 2006) as well as a set of tools to monitor student progress during learning (Dunn & Mulvenon, 2009). In addition, formative assessment often is defined by its purpose or usage, labeling any set of activities or tools as "formative" when the information is used to inform or adapt instruction (Perie, Marion, Gong, & Wurtzel, 2007). Heritage, Kim, Vendlinski, and Herman (2009) define formative assessment as "a systematic process to continuously gather evidence and provide feedback about learning while instruction is under way" (p. 24). Popham (2008) adds a critical clarification: formative assessment is always a planned process; it does not happen accidentally. Taken together, formative assessment is a process in which teachers use various tools and strategies to determine what students know, identify gaps in understanding, and plan future instruction to improve learning.

Applying formative assessment approaches systemically across schools and districts can be a challenge in that they are not based on a single product, technology, or system that can easily be installed or implemented. Changing the relationship between teachers and students and managing these interactions and relationships is at the heart of effective formative assessment. This requires professional development to train, empower, and support the teachers, principals, and district administrators charged with overseeing the quality of such a systemic change. A systemic approach that takes all key stakeholders (students, teachers, and education leaders) into consideration has the best chance for success (Pinchok & Brandt, 2009).

To ensure a cohesive development strategy with alignment of curriculum, instruction, and assessment, it is important to involve content-area experts in the development and implementation processes. Districts should also review their policies and practices to ensure that barriers to progress in formative assessment practices are removed. Long-standing grading and student feedback policies can bring conflict in this realm and can have a negative impact on student learning (Black, Harrison, Lee, Marshall, & William, 2004). Additional time to support formative assessment practices, which are closely aligned with best practices in instructional techniques, should also be allocated.

What was the amount of funding for the program?

Methodology

The amount of funding allotted for the program was obtained from funding workscopes for Title II, Part A. Additional information was collected from Special Revenue Funds Management and individual program managers.

Results

Dallas ISD used $186,357 of Title II grant funds to implement the Assessment for Learning program in 2020-21. The program hired a coordinator to directly support campuses, lead professional development sessions and coach teachers and campus leaders. The remainder of the funds was used toward supplies to support this work.

How was the Assessment for Learning program implemented?

Methodology

Evaluation of the Assessment for Learning program included interviews with program management and a review of internal documents. The purpose of the interviews was to gather information about program requirements and activities. Follow-up information was collected by email.

Results

Program Characteristics

In 2020-21, the Assessment for Learning program took place on eleven campuses: Ben Milam, Kiest, Zaragoza, Stevens Park, Connor, and Tatum Elementary Schools; Franklin and Spence Middle Schools; and Woodrow Wilson, Molina, and Adamson High Schools. A total of 139 teachers, across all grade levels and content areas, implemented the program in their classrooms. Assessment for Learning utilizes an innovative and highly communicative approach to teaching that emphasizes formative assessment throughout instruction.

Formative assessment is a planned, ongoing process used by students and teachers during learning and teaching to elicit and use evidence of student learning to improve student understanding of intended content area learning outcomes and support students to become self-directed learners. Effective use of the formative assessment process requires students and teachers to integrate and embed the following practices in a collaborative and respectful classroom environment: clarifying learning goals and success criteria within a broader progression of learning; eliciting and analyzing evidence of student thinking; engaging in self-assessment and peer feedback; providing actionable feedback; and using evidence and feedback to move learning forward by adjusting learning strategies, goals, or next instructional steps (Fisher, Frey, & Hattie, 2018). Teachers operate in daily feedback loops where students and teachers answer the following questions: Where am I going? Where am I now? How do I close the gap? or Where to next? (Sadler, 1998).

Ten Dimensions of Formative Assessment

Ten dimensions of formative assessment practice have been identified by researchers that can be observed during a lesson (Wylie & Lyon, 2016). The dimensions represent a set of integrated formative assessment practices and form the basis of the Assessment for Learning program.

Learning Goals

Learning Goals should be clearly identified and communicated to students and should help students make connections among lessons within a larger sequence. Learning goals should be aligned to state or district grade-level standards, although this dimension focuses on how the teacher identifies the learning goals for a lesson, communicates them to the students, and uses them in a way that supports learning.

Criteria for Success

Criteria for Success should be clearly identified and communicated to students. This dimension focuses on how the teacher identifies the criteria for success for a lesson and communicates them to the students.

Tasks and Activities to Elicit Evidence of Learning

The focus of this dimension is on those things with which students engage that potentially produce evidence of student learning, excluding classroom discussions. Teachers should use a range of tasks and activities to collect relevant evidence of student thinking. When students are engaged in tasks and activities (on their own, with another student, or in a small group) the work products provide evidence of student understanding. In order to be effective, students should have access to appropriate support from either the teacher or from peers to complete the task. In addition, the teacher needs to have a mechanism for synthesizing evidence from students, whether through a formal review process or informal on-the-fly review.

Questioning Strategies to Elicit Evidence of Learning

The focus of this dimension is on one way that a teacher can collect evidence of student progress through classroom questioning. Teachers should use a range of questioning strategies to collect relevant evidence of student thinking, from more students, more often, and more systematically. Often teachers ask questions only to a few interested students, answer their own questions, or limit student thinking by the type of questions asked. If a teacher has weak questioning strategies, she or he loses opportunities to gain valuable insights into student learning. Teachers can elicit evidence of student thinking through the types of questions students ask of the teacher and peers, as well.

Extending Thinking During Discourse

Students should be provided with ongoing feedback that helps them develop ideas and understanding of the content. This dimension focuses on the teacher's role in providing ongoing feedback during class discussions.

Descriptive Feedback

Students should be provided with evidence-based feedback that is linked to the intended instructional outcomes and criteria for success. This dimension focuses on the teacher's role in providing

individualized feedback to students. This dimension is specific to more formal feedback that tends to be given to individual students on a specific piece of work, either in written form or orally (e.g., during student/teacher conferences) by the teacher.

Peer Feedback

Peer review and feedback is important for providing students an opportunity to think about the work of their peers. This dimension includes the role of student-to-student feedback, while various approaches to teacher feedback are addressed in Feedback Loops and Individualized Descriptive Feedback.

Self-Assessment

"Self-assessment provides students with an opportunity to think meta-cognitively about their learning" (Wylie & Lyon, 2016). An improved understanding of one's own learning is a critical strategy that can lead to improvements in learning.

Collaborative Culture of Learning

A classroom culture in which teachers and students are partners in learning should be established. Classrooms that promote thinking and learning, student autonomy, and students as learning resources for one another may be more successful in encouraging lifelong learners.

Use of Evidence to Inform Instruction

Formative assessment is a process used by teachers and students during instruction that provides feedback to adjust ongoing teaching and learning to improve students' achievement of intended instructional outcomes. This dimension focuses on the teacher use of evidence to adjust instruction across the lesson(s) as a whole.

Campus Walks and Observations

Assessment for Learning program staff utilize the Formative Assessment Rating Rubrics and Observation Protocol (FARROP) to determine implementation of the ten dimensions of formative assessment in classrooms. The complete FARROP rubric can be found in Appendix A. The Assessment for Learning coordinator conducted a total of four campus walks throughout the year (November, February, March, and May). Because the program's results were noted informally and did not involve collecting quantitative measures, an analysis of results was not possible.

In addition, program staff and campus leaders provided coaching and training for teachers in the year three pilot program through observations and POD (Plan, Observe, Debrief) cycles. A POD cycle is a one-and-a-half-day process where teachers plan together, observe each other's classrooms, and receive feedback from one another and their trainer. In previous years, the program conducted formal POD cycles in November, January, and March. However, due to the shortage of substitute teachers in 2020-21, formal POD cycles were impossible. Instead, campuses engaged in "mini-PODs" where they used conference or Professional Learning Community (PLC) time to observe each other. At other campuses, teachers recorded video of lessons and used the POD protocol to give each other feedback after viewing.

Teacher Training

The Assessment for Learning program provided several traditional professional development opportunities throughout the year in addition to the ongoing observation and coaching cycles. A total of 1,137 district teachers, principals, and central staff members attended an online presentation by Doug Fisher on September 1, 2020. A total of 2,532 district teachers, principals, and central staff members attended a professional development session led by formative assessment expert John Hattie on January 4, 2021. Further, 4,236 district teachers, principals, and central staff members attended training with John Almarode on improving distance learning, on March 10, 2021. Assessment for Learning teachers also had access to asynchronous digital training materials throughout the year via Schoology, a learning management system managed by PowerSchool.

Simultaneous Learning

Teachers within the Assessment for Learning program implement simultaneous learning, with distance and in-person students in the same classroom at the same time. Through this approach, the program seeks to create an equitable environment through shared culture, access, and expectations for both in-person and distance learners. Teachers were trained to design the physical classroom environment to support both distance and in-person learners while selecting appropriate videoconferencing tools to leverage screen sharing for equitable access. Teachers established norms and expectations for both cohorts of students without attempting to plan two different courses.

Simultaneous Learning emphasized providing clear learning goals, success criteria, and expectations and building in opportunities for collaboration, feedback, and assessment. Teachers explained the learning goals, success criteria, norms, and expectations. They emphasized assigning student roles to promote agency and collaboration, while building in opportunities for teacher- and student-led small groups. Teachers were encouraged to intentionally create moments of relationship-building within and across modalities of in-person and distance learning.

What were teacher perceptions of the Assessment for Learning program?

Methodology

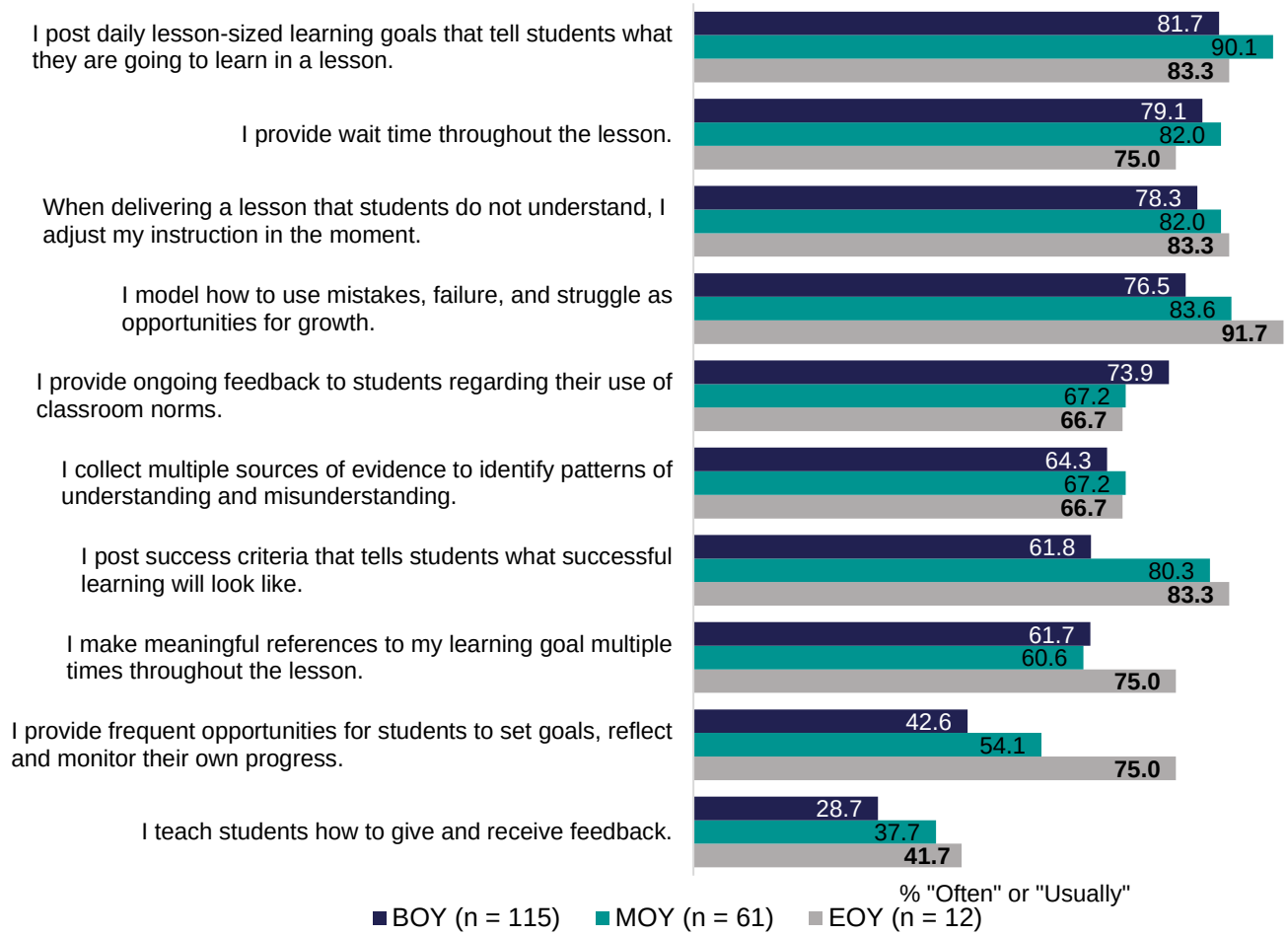
The Assessment for Learning program utilized Google Forms to craft and administer a feedback survey exploring teachers' evolving implementation of Assessment for Learning practices. The program manager provided an online survey link to Assessment for Learning teachers at three different points throughout the year. The beginning-of-year (BOY) survey link was open from September 10, 2020 to November 17, 2020. The middle-of-year (MOY) survey link was open from January 23, 2021 to March 7, 2021. The end-of-year (EOY) survey link was open from May 24, 2021 until June 18, 2021. The purpose of the survey was to collect feedback related to participants' individual experiences working in the Assessment for Learning program. Survey included one demographic item and 10 items with Likert-type response options rating staff perceptions of the program. All Assessment for Learning teachers (n = 124) were selected to receive the survey. Program staff shared raw survey data with the program evaluator. The

evaluator used frequency analyses to determine rates of positive responses for each scaled item, along with number and percentage of responses per category for other items. Results of all survey items, along with a complete distribution of responses is presented in Appendix B

Results

Overall, 115 respondents completed at least one survey, resulting in a 93 percent response rate, though participation dropped sharply as the year progressed (61 teachers responded during the MOY survey window, and 12 teachers responded during the EOY survey window.) Figure 1 shows the percentage of respondents who indicated that they "usually" or "often" did each selected item.

Figure 1: 2021 Assessment for Learning Teacher Survey Responses



Source: Assessment for Learning teacher survey results files dated 06/10/21 and 06/21/21.

Overall, the proportion of teachers who indicated that they incorporated each behavior into their teaching increased over the course of the school year. The largest increases were seen in posting success criteria that tells students what successful learning will look like (61.8% at BOY versus 83.3% at EOY, an increase of 21.5 percentage points) and providing opportunities for students set goals, reflect, and monitor

their own progress (42.6% at BOY versus 75.0% at EOY, an increase of 32.4 percentage points). At the beginning of the year, 76.5 percent of teachers reported that they modeled how to use mistakes, failure, and struggle as opportunities for growth. By the end of the year, 91.7 percent of teachers reported "often" or "usually" doing so. The formative assessment skill rated least implemented at the beginning of the year (28.7% at BOY), teaching students how to give and receive feedback, increased by 13 percentage points by the end of the year (41.7% at EOY).

How did participation in the Assessment for Learning program affect student achievement?

Methodology

The NWEA *Measures of Academic Progress* (*MAP)* assessments are universal screeners administered to Dallas ISD students in grades one through ten at the beginning (BOY), middle (MOY), and end (EOY) of the school year. *MAP* data consists of several variables used to describe student performance and academic growth over time by subject. The Rasch Unit (RIT) score represents the level at which students are ready to learn at the time of assessment. The RIT score is a standardized, equal-interval score that can be compared to other RIT scores regardless of student age or grade. The second *MAP* Growth variable used for this report was the Conditional Growth Percentile (CGP). The CGP represents the extent to which student achievement changed from one testing window to the next, relative to change for other students who are similar in starting achievement level, grade, and number of instructional weeks between testing windows. For each subject assessed, CGPs for each student are calculated with respect to a large, nationally-normed sample obtained by NWEA and can be calculated whenever a student takes a *MAP* assessment in a new testing window (e.g., BOY to MOY, MOY to EOY, BOY to EOY). A CGP at the 50th percentile indicates that the RIT score increase from the prior testing window represents the median amount, or expected level, of progress for that grade level[3].

The evaluator used student-level *MAP* data files dated November 5, 2020 (BOY), March 30, 2021 (MOY), and June 23, 2021 (EOY) and merged them with the district demographic data files pulled closest to the dates of the *MAP* Growth data files for each testing window. The evaluator selected a sample of 50 grades one through 11 Assessment for Learning teachers. Students were flagged as participants if they were identified by district course files as being taught by one of these teachers. To construct comparison groups of comparable classrooms not implementing the program, classrooms of the same grade on the same campus as implementing teachers were chosen whenever possible. In the case of Ben Milam Elementary, where the entire campus was implementing the program, eight teachers from the same grade levels at Lipscomb Elementary were selected. This campus was chosen based on similarity of campus-level demographic and achievement criteria used to create state accountability comparison groups. Students were selected into the sample if they had valid *MAP* scores for all three test administrations (English

[3] For more information, see https://community.nwea.org/docs/DOC-1630.

language only in Reading, as English and Spanish *MAP* Reading assessments are not directly comparable). In the cases where students appeared on rosters for both AfL and non-AfL teachers (such as high school students taking courses from multiple teachers on a campus), the student was omitted from the study. The evaluator then used frequency and crosstabs analyses to calculate the average *MAP* RIT and CGP scores by subject, grade, and student group. The evaluator conducted *t*-tests of mean RIT scores to ascertain if the program made a statistically significant difference in the participants' academic growth as measured by *MAP* results. Also, the evaluator used *Cohen's d* to examine if differences were practically meaningful. Finally, the evaluator conducted chi-square tests of independence to determine if a significant association existed between AfL participation and growth threshold rates as illustrated by CGP score.

Results

BOY, MOY, and EOY *MAP* RIT data by grade level are shown in Tables 1 and 2 (reading) and Tables 3 and 4 (mathematics). At all grade levels, RIT scores for reading and mathematics generally remained unchanged or increased from BOY to MOY and from MOY to EOY. For each assessment and testing period, mean RIT scores generally increased for higher grade levels, as expected due to the nature of the assessment (scores are given on a single continuum of RIT scores from grades one through 11).

While AfL students' overall mean Reading RIT across all grades was five points higher than non-AfL students' at BOY (210 versus 205; Table 1 and Table 2, respectively), the mean RIT for Afl students increased by three points by MOY (210 to 213), while increasing by four points for non-AfL students (205 to 209). From MOY to EOY, AfL students across all grades lost the previous gains, with an overall mean RIT score of 210, the same as at BOY. For non-AfL students, the overall mean EOY RIT score was three points higher than at BOY (205 versus 208). Notably, students at grades one through three saw larger gains from BOY to EOY than upper grades: 11 to 18 points for AfL students and nine to 15 points for non-AfL students. In addition, AfL students at grade eight saw a loss of six points (206 to 200) from BOY to EOY, while non-AfL students gained one point during this time period (227 to 228).

While there were slight differences in mean RIT scores, t-test and effect size analyses of the differences between AfL and non-AfL students' mean RIT Reading scores at BOY, MOY, and EOY showed no statistical or practical significance. For BOY, $t(20) = 0.17$, $p = 0.86$[4], *Cohen's d* = 0.19[5]. For MOY, $t(20) = 0.22$, $p = 0.82$, *Cohen's d* = 0.18. For EOY, $t(20) = 0.28$, $p = 0.78$, *Cohen's d* = 0.10.

[4] The smaller the *p*-value, the stronger the evidence that you should reject the null hypothesis. A *p*-value less than 0.05 (typically ≤ 0.05) is statistically significant. A *p*-value higher than 0.05 (> 0.05) is not statistically significant and indicates strong evidence for the null hypothesis (McLeod, 2019).

[5] Cohen's *d* values of .20, .50, and .80 are considered small, medium, and large, respectively (Cohen, 1992).

Table 1: 2020-21 *MAP* English Reading Assessment Mean RIT Scores by Grade, Students of Teachers Implementing Assessment for Learning

Grade Level	*n*	BOY Mean RIT	MOY Mean RIT	EOY Mean RIT	BOY to EOY *Diff.*
1	15	158	171	176	18
2	*	*	*	*	*
3	50	183	192	194	11
4	75	189	194	194	5
5	19	208	210	212	5
6	52	212	213	212	0
7	44	209	212	210	1
8	10	206	207	200	-6
9	203	223	226	220	-3
10	66	221	222	217	-4
11	25	216	221	214	-2
Overall	**561**	**210**	**213**	**210**	**0**

Source: District BOY, MOY, and EOY *MAP* data files dated 11/05/20, 03/30/21, and 06/23/21, respectively. District course data files dated 10/29/20 (BOY), 03/04/21 (MOY), and 06/02/21 (EOY). Assessment for Learning teacher identification file dated 06/10/21.
Note: Includes data for students with valid RIT scores for the English reading assessment.
"*" Indicates results are masked due to small numbers to protect student confidentiality.

Table 2: 2020-21 *MAP* English Reading Assessment Mean RIT Scores by Grade, Comparison Group of Students of Teachers Not Implementing Assessment for Learning

Grade Level	*n*	BOY Mean RIT	MOY Mean RIT	EOY Mean RIT	BOY to EOY *Diff.*
1	28	163	169	176	13
2	42	171	179	186	15
3	31	186	193	195	9
4	44	191	194	194	3
5	38	199	203	205	3
6	57	207	214	211	4
7	76	212	217	212	0
8	68	227	232	228	1
9	117	214	215	214	0
10	43	215	217	213	-2
11	40	212	212	209	-3
Overall	**584**	**205**	**209**	**208**	**3**

Source: District BOY, MOY, and EOY *MAP* data files dated 11/05/20, 03/30/21, and 06/23/21, respectively. District course data files dated 10/29/20 (BOY), 03/04/21 (MOY), and 06/02/21 (EOY). Assessment for Learning teacher identification file dated 06/10/21.
Note: Includes data for students with valid RIT scores for the English reading assessment.

Overall mean Mathematics RIT scores for AfL students was three points higher at BOY than for non-AfL students (212 versus 209; Table 3 and Table 4, respectively). From BOY to MOY, the mean RIT score increased by eight points for both AfL students (212 to 220) and non-AfL students (209 to 217). From MOY to EOY, AfL students across all grades gained two points (220 to 222). For non-AfL students, the mean RIT score increased by three points from MOY to EOY (217 to 220). Again notably, students at lower grades, this time one through three, saw notable gains from BOY to EOY, 15 to 23 points for AfL students and 11 to 16 points for non-AfL students. Again, *t*-test and effect size analyses of the differences between AfL and non-AfL students' mean RIT Mathematics scores at BOY, MOY, and EOY showed no statistical or

practical significance. For BOY, $t(20) = 0.37$, $p = 0.72$, *Cohen's d* = 0.13. For MOY, **t**(20) = 0.28, $p = 0.78$, *Cohen's d* = 0.09. For EOY, $t(20) = 0.32$, $p = 0.76$, *Cohen's d* = 0.08.

Table 3: 2020-21 *MAP* Mathematics Assessment Mean RIT Scores by Grade, Students of Teachers Implementing Assessment for Learning

Grade Level	*n*	BOY Mean RIT	MOY Mean RIT	EOY Mean RIT	BOY to EOY *Diff.*
1	15	158	179	181	23
2	*	*	*	*	*
3	50	181	193	196	15
4	75	190	197	202	12
5	19	203	212	215	12
6	52	211	218	221	10
7	44	216	222	224	8
8	10	204	215	216	12
9	203	228	234	236	8
10	66	224	231	232	8
11	25	225	232	230	5
Overall	**561**	**212**	**220**	**222**	**10**

Source: District BOY, MOY, and EOY *MAP* data files dated 11/05/20, 03/30/21, and 06/23/21, respectively. District course data files dated 10/29/20 (BOY), 03/04/21 (MOY), and 06/02/21 (EOY). Assessment for Learning teacher identification file dated 06/10/21.
Note: Includes data for students with valid RIT scores for the mathematics assessment.
"*" Indicates results are masked due to small numbers to protect student confidentiality.

Table 4: 2020-21 *MAP* Mathematics Assessment Mean RIT Scores by Grade, Students of Teachers Not Implementing Assessment for Learning

Grade Level	*n*	BOY Mean RIT	MOY Mean RIT	EOY Mean RIT	BOY to EOY *Diff.*
1	28	167	175	182	15
2	42	174	182	185	11
3	31	187	197	203	16
4	44	195	204	204	9
5	38	200	209	212	12
6	57	210	218	224	14
7	76	216	223	224	8
8	68	235	246	249	14
9	117	218	226	228	10
10	43	219	228	230	11
11	40	219	225	226	7
Overall	**584**	**209**	**217**	**220**	**11**

Source: District BOY, MOY, and EOY *MAP* data files dated 11/05/20, 03/30/21, and 06/23/21, respectively. District course data files dated 10/29/20 (BOY), 03/04/21 (MOY), and 06/02/21 (EOY). Assessment for Learning teacher identification file dated 06/10/21.
Note: Includes data for students with valid RIT scores for the mathematics assessment.

Table 5 shows *MAP* Reading and Mathematics growth thresholds based on student CGP by comparison group (all grades included) for growth demonstrated between the BOY and MOY assessments. To measure growth across testing periods, the evaluator divided CGP scores into three categories based on NWEA's definition of meeting expected growth. Students with a CGP below the 45th percentile "failed to meet" nationally-normed growth expectations. Those with a CGP between the 46th and 55th percentile

"met" expectations, and those with a CGP above the 55th percentile "exceeded" those expectations for growth.

For reading, the percentage of AfL students who met or exceeded growth thresholds by MOY was 55.4 percent which was lower by 2.5 percentage points compared to Non-AfL students (57.9%). For mathematics, the percentage of AfL students who met or exceeded growth thresholds by MOY was 71.1 percent which was higher by 1.7 percentage points compared to Non-AfL students (69.4%). There were no statistically significant differences between students of AfL and Non-Afl teachers in meeting or exceeding growth expectations (2.5 percentage points or less). Further, A chi-square test of independence showed that there was no significant association between AfL participation and growth threshold rates in either Reading (X^2 = 5.03, p = .08) or Mathematics (X^2 = 0.49, p = .78).

Table 5: 2020-21 Dallas ISD *MAP* Reading (English Only) and Mathematics Growth Threshold Rates by Assessment for Learning (AfL) Comparison Group: BOY to MOY

	Reading						Mathematics					
	Failed to Meet		Met		Exceeded		Failed to Meet		Met		Exceeded	
Group	***n***	**%**	***n***	**%**	***n***	**%**	***N***	**%**	***n***	**%**	***n***	**%**
AfL	250	44.6	46	8.2	265	47.2	162	28.9	27	4.8	372	66.3
Non-AfL	246	42.1	32	5.5	306	52.4	179	30.5	30	5.1	377	64.3
Overall	**496**	**43.3**	**78**	**6.8**	**571**	**49.9**	**341**	**29.7**	**57**	**5.0**	**749**	**65.3**

Source: District BOY and MOY *MAP* data files dated 11/05/20, and 03/30/21, respectively. District course data files dated 10/29/20 (BOY) and 03/04/21 (MOY). Assessment for Learning teacher identification file dated 06/10/21.
Note: Includes data for students who had valid *MAP* scores (English language only for Reading). Failed to Meet: student conditional growth percentile (CGP) < 45th percentile. Met: 45th percentile <= CGP >=55th percentile. Exceeded: CGP > 55th percentile. For each assessment and group, percentages of students in the Failed to Meet vs. Met vs. Exceeded categories may not sum to 100 percent because of rounding.

Table 6 shows *MAP* Reading and Mathematics growth thresholds based on student CGP by comparison group (all grades included) for growth demonstrated between the MOY and EOY assessment. For reading, the percentage of AfL students who met or exceeded growth thresholds by EOY was 34.2 percent which was lower by 4.8 percentage points compared to Non-AfL students (39.0%). For mathematics, the percentage of AfL students who met or exceeded growth thresholds by EOY was 49.4 percent which was lower by 1.5 percentage points compared to Non-AfL students (50.9%). There were no statistically significant differences between students of AfL and Non-Afl teachers in meeting or exceeding growth expectations (4.8 percentage points or less). Further, A chi-square test of independence showed that there was no significant association between AfL participation and growth threshold rates in either Reading (X^2 = 5.03, p = .08) or Mathematics (X^2 = 0.26, p = .88).

Table 6: 2020-21 Dallas ISD *MAP* Reading (English Only) and Mathematics Growth Threshold Rates by Assessment for Learning (AfL) Comparison Group: MOY to EOY

	Reading						Mathematics					
	Failed to Meet		Met		Exceeded		Failed to Meet		Met		Exceeded	
Group	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
AfL	369	65.8	27	4.8	165	29.4	284	50.6	28	5.0	249	44.4
Non-AfL	356	61.0	28	4.8	200	34.2	288	49.1	31	5.3	267	45.6
Overall	**725**	**63.3**	**55**	**6.1**	**365**	**31.9**	**572**	**49.9**	**59**	**5.1**	**516**	**45.0**

Source: District MOY and EOY *MAP* data files dated 03/30/21 and 06/23/21, respectively. District course data files dated 03/04/21 (MOY), and 06/02/21 (EOY). Assessment for Learning teacher identification file dated 06/10/21.
Note: Includes data for students who had valid *MAP* scores (English language only for Reading). Failed to Meet: student conditional growth percentile (CGP) < 45th percentile. Met: 45th percentile <= CGP >=55th percentile. Exceeded: CGP > 55th percentile. For each assessment and group, percentages of students in the Failed to Meet vs. Met vs. Exceeded categories may not sum to 100 percent because of rounding.

What are the plans for the future of the Assessment for Learning program?

Methodology

Evaluation of the Assessment for Learning program included interviews with program management and a review of internal documents. The purpose of the interviews was to gather information about program requirements and activities. Follow-up information was collected by email.

Results

Planning for 2021-22

The Assessment for Learning Design Team was an advisory group consisting of program management, representatives from the Teaching and Learning and Evaluation and Assessment departments, AfL campus principals, and AfL teachers. The Design Team met four times during the year, in November, December, February, and June, providing feedback on program implementation, sharing campus and classroom challenges and solutions, and advising program management on program expansion plans.

On Saturday June 12, 2021, the program held an Assessment for Learning Summit on Zoom. The event lasted from 9:00 am to 1:00 pm and was open to all AfL teachers and principals. Three guest presenters joined district staff for author- and teacher-led breakout rooms, and discussions around books supplied to attendees upon registration. Formative assessment expert John Almarode addressed learning goals, success criteria, and feedback. Author Anthony Muhammad discussed equity and transforming school culture. Author and educational consultant Dr. Nicole Law presented on culturally responsive pedagogy and visible learning. A total of 198 district staff members attended.

While the program was implemented by at least some teachers at 11 campuses in 2020-21, the district is seeking to expand the program to include additional high priority campuses in 2021-22. Additionally, the program will expand to campus-wide implementation at several existing AfL campuses including Adamson and Wilson High Schools, Spence Middle School, and Conner, Stevens Park, Tatum,

and Zaragoza Elementary Schools. Program management and district leadership are closely monitoring data to see if these practices could potentially be scaled district-wide as early as 2022-23.

KEY FINDINGS AND RECOMMENDATIONS

Key Findings

The Assessment for Learning program sought to shift the thinking and practices around formative assessment in Dallas ISD by training teachers to intentionally incorporate formative assessment into their daily instruction with the support from coaching and professional development provided by program staff and experts in the field. In 2020-21, the district scaled up the program to include teachers at eleven campuses in the professional development and coaching process and continued to train leaders to better utilize formative assessment. Dallas ISD used $186,357 of Title II grant funds to implement the Assessment for Learning program in 2020-21. The program hired a coordinator to directly support campuses, lead professional development sessions and coach teachers and campus leaders.

- The Assessment for Learning coordinator conducted a total of four campus walks throughout the year (November, February, March, and May). Because the program's results were noted informally and did not involve collecting quantitative measures, an analysis of results was not possible.
- Due to the shortage of substitute teachers in 2020-21, formal POD cycles were impossible. Instead, campuses engaged in "mini-PODs" where they used conference or Professional Learning Community (PLC) time to observe each other. At other campuses, teachers recorded video of lessons and used the POD protocol to give each other feedback after viewing.
- The Assessment for Learning program provided several traditional professional development opportunities throughout the year in addition to the ongoing observation and coaching cycles. Assessment for Learning teachers also had access to asynchronous digital training materials.
- Overall, the proportion of teachers who indicated that they incorporated key formative assessment behavior into their teaching increased over the course of the school year, as indicated through the AfL Teacher Survey. The largest increases were seen in posting success criteria that tells students what successful learning will look like (61.8% at BOY versus 83.3% at EOY, an increase of 21.5 percentage points) and providing opportunities for students set goals, reflect, and monitor their own progress (42.6% at BOY versus 75.0% at EOY, an increase of 32.4 percentage points).
- While differences in mean RIT scores from BOY, MOY, and EOY administrations of *MAP* Reading and Mathematics assessments were observed, t-test and effect size analyses of the differences between AfL and non-AfL students' scores showed no statistical or practical significance.
- There were no statistically significant differences (4.8 percentage points or less in Reading and 1.5 percentage points or less in Mathematics) between students of AfL and Non-Afl teachers in meeting or exceeding growth expectations between MOY and EOY.
- While the program was implemented by at least some teachers at 11 campuses in 2020-21, the district is seeking to expand the program to include additional high priority campuses in 2021-22.

Additionally, the program will expand to campus-wide implementation at several existing AfL campuses.

Recommendations

Based on key findings, the following recommendations are offered for stakeholders.

Formalize campus walk feedback. Currently, campus walk data is captured through non-standardized observer notetaking. Formalizing the campus walk feedback process to include an observation checklist would allow for quantitative analysis of the quality of instruction observed, including tracking change over time.

Continue strong collaboration practices, including design team, PODs, and campus walks. As previously discussed, applying new formative assessment approaches across schools is challenging because it aims to fundamentally change the relationship between teachers and students. A formal approach to collaboration and sharing of best practices between teachers and campus administrators increases the chance for success.

REFERENCES

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. Phi Delta Kappan, 86(1), 8–21.

Cohen, J. (1992). Fuzzy methodology. *Psychological Bulletin, 112*(3), 409–410.

Council of Chief State School Officers. (2007). Formative assessment and CCSSO: A special initiative—a special opportunity. Washington, DC: Author.

Dunn, K. E., & Mulvenon, S. W. (2009). A critical review of research on formative assessment: The limited scientific evidence of the impact of formative assessment in education. Practical Assessment, Research & Evaluation, 14(7), 1–11.

Fisher, D., Frey, N., Hattie, J. (2018). Developing Assessment-Capable Visible Learners, Grades K-12: Maximizing Skill, Will, and Thrill. United States: SAGE Publications.

Heritage, M., Kim, J., Vendlinski, T., & Herman, J. (2009). From evidence to action: A seamless process in formative assessment? Educational Measurement: Issues and Practice, 28(3), 24–31.

McLeod, S. A. (2019, May). *What a p-value tells you about statistical significance*. Simply Psychology. https://www.simplypsychology.org/p-value.html

Perie, M., Marion, S., Gong, B., & Wurtzel, J. (2007). The role of interim assessments in a comprehensive assessment system. Washington, DC: The Aspen Institute.

Pinchok, N., & Brandt, W.C. (2009). Connecting Formative Assessment Research to Practice: An Introductory Guide for Educators. Naperville, IL: Learning Point Associates.

Popham, W. J. (2006). Phony formative assessments: Buyer beware! Educational Leadership, 64(3), 86–87.

Popham, W. J. (2008). Transformative assessment. Alexandria, VA: ASCD.

Sadler, D. Royce (1998) Formative Assessment: Revisiting the Territory, *Assessment in Education: Principles, Policy & Practice*, 5(1), 77-84.

Wylie, C., & Lyon, C. (2016, February). *Using the Formative Assessment Rubrics, Reflection and Observation Tools to Support Professional Reflection on Practice*. Educational Testing Services. https://center.ncsu.edu/ncfalcon/pluginfile.php/2/course/section/57/FAROP_Revised_2016.pdf

APPENDICES

Appendix A: Complete FARROP Rubric

Appendix A, Table 1: Learning Goals

1 - Beginning	2 - Developing	3 - Progressing	4 - Extending
☐ The focus of the lesson is presented in isolation and without connecting toprevious learning, to future learning, or to a broader purpose for the learning. *OR* Superficial procedural connections are made (e.g., "We started argumentation yesterday" or "We'll wrap up problem-solving strategies tomorrow"), or a topic is identified without providing specific goals. *OR* The content of the learninggoals is highly inappropriate for the students. *OR* The learning goals are expressed in language thatis not accessible to students.	☐ The focus of the lesson is presented with only isolated references made to previous learning, to future learning, orto a broader purpose for the learning.	☐ The focus of the lesson is clearly presented in terms of previous or future learning. A larger sequence of learning is identified, andthe teacher explains how the current lesson fits within the larger sequence or how it contributes to broader purpose for the learning.	☐ The focus of the lesson is presented as part of a coherent sequence of learning, with meaningful connections made to previous orfuture learning in a way that facilitates students' clear understanding of the connections or in a way that contributes to abroader purpose for the learning.
	☐ The learning goals focus on what students should know, understand, or be able to do by the end of the lesson. The content of the learning goals is appropriate for students and is expressed in language that is accessible to students,but opportunities for students to internalize the learning goals are not provided.	☐ The learning goals focus on what students should know, understand, or be able to do bythe end of the lesson. The content of the learning goals is appropriate for students andis expressed in language that is accessible to students, and opportunities for students to internalize the learning goals are provided.	☐ The learning goals focus on what students should know, understand, or be able to do by the end of the lesson. The content of the learning goals is appropriate for students and is expressed in language that is accessible to students; opportunities for students to internalize the learning goals are provided; andthe teacher checks for understanding.
	☐ The teacher presents the learning goals to students but makes no verbal or directreference to the learning goals near the start of the lesson.	☐ The teacher presents the learning goals to students and makes verbal or direct referenceto the learning goals near the start of the lesson.	☐ The teacher presents the learning goals to students and makes meaningful and appropriate reference to the learning goals atthe start of the lesson.
	☐ The teacher does not return to the learning goals at any point during thelesson.	☐ The teacher makes some reference back to the learning goals toward the end of the lesson, in a way that superficially focuses student attention on the purpose of the lesson.	☐ The teacher makes multiple meaningful and appropriate verbal references to the learning goals throughout the lesson, summarizes progress toward the goals near the end of the lesson in ways that support student learning, orinvites students to explain the learning goals atthe end of the lesson.

Appendix A, Table 2: Criteria for Success

<table>
<tr><th>1 - Beginning</th><th>2 - Developing</th><th>3 - Progressing</th><th>4 - Extending</th></tr>
<tr><td rowspan="2">☐ The criteria for success are not appropriate for the learning goals (e.g., they only refer to task requirements rather than helping students understand what quality work would look like in relation tothe learning goals) or are not appropriate for students.
OR
The criteria for success are expressed in language that is not accessible to students.
OR
The teacher makes only a reference tocriteria, such as “I can” statements, but without any explanation or presentation (e.g., “When you are done with the problem, you will usethe rubric to score it”), and studentsdo not seem to be familiar with the rubric and/or are not able to use it meaningfully.</td><td>☐ The criteria for success are somewhat appropriate for the learning goals and for students. They include extraneous task information and/or do not refer tokey elements of the learning goal. They are generally expressed in language that is accessible to the students.</td><td>☐ The criteria for success are appropriate for the learning goals and for students, and they are expressed in language that is accessible to the students.</td><td>☐ The criteria for success are appropriate for the learning goals and for students. They clearly articulate what quality work looks likeand are expressed in language that is accessible to the students.</td></tr>
<tr><td>☐ The teacher presents or reviews thecriteria with students but does not provide a way for students to internalize the criteria or to use the criteria effectively, resulting in few students engaging with the criteria in meaningful ways.</td><td>☐ The teacher engages the students with the criteria by providing a way for students to internalize the criteriaand/or use the criteria effectively, but only some students seem to understand or engage with the process in meaningful ways.</td><td>☐ The teacher deeply engages the students with the criteria by providing a way for students to internalize the criteria and/or usethe criteria effectively, allowing the majorityof students to engage with the criteria in meaningful ways that support learning throughout the lesson.</td></tr>
</table>

Appendix A, Table 3: Questioning Strategies that Elicit Evidence of Student Learning

1 - Beginning	2 - Developing	3 - Progressing	4 - Extending
☐ The teacher asks very few questions designed to elicit evidence of the learning goalsand to encourage discourse during the lesson.	☐ The teacher asks questions designed to elicit evidence of the learning goals andto encourage classroom discourse at a few points during the lesson, or the teacher asks questions that are not integrated into instruction.	☐ The teacher asks questions designed to elicit evidence of the learning goals and to encourage classroom discourse periodically; or the teacher asks questionsmore frequently, but the questions are not well integrated into instruction.	☐ Throughout the lesson, the teacher asks questions designed to elicit evidence of the learning goals and to encourage classroom discourse; questioning and discussion are seamlessly integrated into instruction.
☐ The teacher provides inadequate wait time and/or often answers his or her own questions.	☐ The teacher infrequently provides adequate wait time. The teacher sometimes answers his or her own questions before students have a chance to respond or even after a student has provided an answer.	☐ The teacher often provides sufficient wait time. The teacher does not answer hisor her own questions before students havea chance to respond or after a student has provided an answer.	☐ The teacher provides sufficient wait time throughout the lesson. The teacherdoes not answer his or her own questions before students have a chance to respond or after a student has provided an answer.
☐ The teacher uses questioning strategies that provide evidence from only a few students or from the same students in the class.	☐ The teacher infrequently uses questioning strategies to collect evidence of learning from a broad sample of students and may implementthem in a way that does not support active engagement from most students.	☐ The teacher often uses effective questioning strategies to collect sufficientevidence of learning from all students in systematic ways and in a way that supports active engagement from most students.	☐ The teacher uses effective questioning strategies to collect evidence of learning from all students in systematic ways and in a way that supports active engagement from most or all students.
☐ The evidence collected cannot be used to make meaningful inferences about the class's progress on intended learning outcomes and to adapt/continue instruction.	☐ There is some evidence that the teacher occasionally capitalizes on opportunitiesto make inferences about student progress and/ or to adapt/continue instruction accordingly.	☐ There is clear evidence that the teacher capitalizes on most opportunities to make inferences about student progress and to adapt/ continue instruction accordingly.	☐ There is strong evidence that the teacher effectively uses student responses and student questions to makeinferences about student progress and to adjust/continue instruction accordingly throughout the lesson.

Appendix A, Table 4: Tasks and Activities that Elicit Evidence of Student Learning

1 - Beginning	2 - Developing	3 - Progressing	4 - Extending
☐ The teacher uses tasks or activities that are not aligned tothe learning goals or will not provide evidence of student progress toward those goals.	☐ The teacher uses tasks or activities that are loosely alignedto the learning goals and will provide limited evidence of student progress toward those goals.	☐ The teacher uses well-crafted tasks and activities that are mostly aligned tothe learning goals and will provide evidence of student progress toward those goals.	☐ The teacher uses a series of integrated, well-crafted tasks and activities that are tightly aligned to the learning goals and will provide evidence of student progress toward those goals.
☐ Most students are unclear about how they need to approachthe task, and students require extensive repeated or revised explanations.	☐ Many students are unclear about how they need to approach the task, and the teacher takes some time to repeat or revise explanations.	☐ A few students are unclear about how they need to approach the task, and the teacher takes minimal time isto repeat or revise explanations.	☐ Most or all students are clear about how theyneed to approach the task and are able to begin work efficiently.
☐ The teacher does not review student work products during thelesson or does not indicate whenthey will be reviewed.	☐ The teacher occasionally or haphazardly reviews student workproducts during the lesson or makes a vague reference to when they will be reviewed.	☐ The teacher reviews student work products during the lesson in a way that provides insight into most students' progress or indicates how work products will be reviewed later.	☐ The teacher systematically reviews student work products during the lesson in a way that provides insight into most or all students' progress or clearly indicates how they will be reviewed and how the information will be used toinform instruction.

Appendix A, Table 5: Using Evidence to Inform Ongoing Teaching & Learning

1 - Beginning	2 - Developing	3 - Progressing	4 - Extending
☐ There is little attempt by the teacher to collect evidence ofstudent learning in the lesson that is aligned to the learning goals or criteria for success. *OR* The collection of evidence is so minimal or inconsistent that there is no way for the teacher togain insight into student learning.	☐ There is some evidence that the teachercollects evidence of student learning thatis somewhat aligned to the learning goalsor criteria for success, but not directly representative of those goals or success criteria.	☐ There is some evidence that the teacher collects evidence of student learning that is aligned to the learning goals or criteria for success throughoutthe lesson.	☐ There are multiple sources of evidence that indicate that the teacher skillfully and systematically collects evidence of student learning that is aligned to the learning goals or criteria for success throughout thelesson.
☐ The teacher does not have evidence of student learning toanalyze	☐ The teacher does not analyze the evidence to identify patterns of understanding/ misunderstanding or to make inferences about student strengths and weaknesses.	☐ There is some evidence that the teacher is analyzing the evidence to identify patterns of understanding/ misunderstanding or to make inferences about student strengths and weaknesses.	☐ There are multiple sources of evidencethat indicate the teacher is analyzing the evidence to identify patterns of understanding/ misunderstanding and to make inferences about student strengths and weaknesses.
☐ The teacher has no basis for modifying instructional plans.	☐ There are no teacher comments that provide any evidence to suggest that student work is used to shape instructional decisions (observable evidence for this level is characterized by lost opportunities).	☐ Teacher comments provide some evidence that the student work, identified patterns, and inferences are used to shape instructional decisions.	☐ Multiple teacher comments provide clear evidence that the student work, identified patterns, and inferences are usedto shape instructional decisions and advance student learning.

Appendix A, Table 6: Descriptive Teacher Feedback

1 - Beginning	2 - Developing	3 - Progressing	4 - Extending
☐ The teacher provides evaluative feedback on a specific piece of work (e.g., a score, grade, or other summative feedback).	☐ The teacher provides feedback on a specific piece of work that is overly general or focused on the student, e.g.,"You did a great job!"	☐ The teacher provides descriptive feedback on a specific piece of work thatis mostly aligned to the learning goals and the criteria for success, though may also contain some general or student-focused comments.	☐ The teacher provides descriptive feedback on a specific piece of work that clearly supports the learning goals and strongly reflects the criteria for success.
☐ Corrective feedback does all thethinking for the students; subsequent student actions consist solely of following directions.	☐ Corrective feedback lets students do some of the thinking but does not appropriately scaffold the next steps thatstudents are to take, e.g., Your conclusion needs work. How could you improve it?	☐ Corrective feedback lets students do the thinking and scaffolds the next stepsstudents are to take by providing a suggestion, reminder, or question to elicit further learning from the students.	☐ Corrective feedback lets students do the thinking, pointing out one or more areas to work on, followed by a suggestion, reminder, or question to elicit further learning from the students. Both the teacher and student are clear about next steps the student can take independently to advance learning.
☐ The teacher does not have routines or structures in place toprovide feedback to most or all students.	☐ The teacher has minimal routines and structures in place for providing feedback to most or all students.	☐ The teacher has adequate routines and structures in place for providing feedback to most or all students.	☐ The teacher has a clear and systematic approach for providing feedback to most orall students.
☐ There is no opportunity for students to review the feedback, ask questions in order to internalizethe feedback, or apply the feedbackto their work in meaningful ways.	☐ There is little or no opportunity for students to review the feedback, ask questions in order to internalize the feedback, or apply the feedback to theirwork in meaningful ways.	☐ Students are provided with limited structures and supports (e.g., limited time is provided, or students are confused about the process) to reviewthe feedback, ask questions in order tointernalize the feedback, or apply the feedback to their work in meaningful ways.	☐ Students are provided with ample structures and supports (e.g., time, feedbackstructures, etc.) to review the feedback, ask questions in order to internalize the feedback, or apply the feedback to their work in meaningful ways.

Appendix A, Table 7: Collaborative Culture of Learning

1 - Beginning	2 - Developing	3 - Progressing	4 - Extending
☐ The classroom climate is characterized by an overall perception that the teacher is in charge of the learning.	☐ The classroom climate is characterizedfor the most part by an overall perception that the teacher is in charge ofthe learning.	☐ The classroom climate is characterized for the most part by an overall perception that the teacher and students are equally responsible for thelearning.	☐ The classroom climate is characterized by an overall, consistent perception that the teacherand students are equally responsible for the learning.
☐ Student-to-student collaboration is not evident.	☐ Minimal student-to-student collaboration is evident.	☐ Some student-to-student collaboration is evident.	☐ Student-to-student collaboration is evident and spontaneous or a preference of the students when theyare given a choice.
☐ Student participation is limited to when the teacher asks a question, and the teacher does not capitalize on student responses or student questions todeepen learning.	☐ Student participation is limited to when the teacher asks a question, and the teacher rarely capitalizes on student responses or student questions to deepenlearning.	☐ Student participation is encouraged, and the teacher often capitalizes on student responses or student questionsto deepen learning.	☐ Student participation is spontaneous and respectful, and the teacher often capitalizes on student responses or student questions to deepen learning.
☐ Multiple viewpoints or approaches arenot sought.	☐ Multiple viewpoints or approaches arerarely sought.	☐ Multiple viewpoints or approachesare occasionally sought.	☐ Multiple viewpoints or approachesare sought throughout the lesson.
☐ The teacher does not demonstrate a growth mindset through comments andquestions.	☐ The teacher does not demonstrate a growth mindset through comments and questions, or the teacher is not convincing.	☐ For the most part, the teacher demonstrates a growth mindset through comments and questions.	☐ Throughout the lesson, the teacher and students demonstrate a growth mindset through their comments and questions.

Appendix A, Table 8: Extended Thinking During Discourse

1 - Beginning	2 - Developing	3 - Progressing	4 - Extending
☐ The teacher asks questions from students, but neither the teacher nor the students build on responses. Rather, discourse focuses on a statement of correct or incorrect rather than deeper/meaningful exploration ofideas.	☐ The teacher and some of the students occasionally build on student responses, or the teacher occasionally encourages studentsto build on each other's responses.	☐ The teacher and some of the students frequently build on other students' responses by clarifying student comments, providing feedback, pushingfor more elaborate answers, or engagingmore students in thinking about the problem. Students sometimes direct questions to each other and respond to other students' questions or statements without prompting.	☐ The teacher and some of the students frequently build on other students' responsesby clarifying student comments, providing feedback, pushing for more elaborate answers, or engaging more students in thinking about the problem. Students ask probing questions of the teacher and of each other during discussions. They often respondto each other's questions or statements without prompting.
	☐ There are occasional feedback opportunities that engage studentsin deepening the discussion, although they are short, often end abruptly, and do not allow a full exploration of ideas and concepts or do not help to develop ideas and/or understanding of the content.	☐ There are multiple feedback opportunities that engage students in deepening the discussion, rarely end withthe teacher indicating correct or incorrectresponses and allow for deeper/more meaningful exploration of some ideas.	☐ There are continuous feedback opportunities that engage students in deepening the discussion through the use of probing questions to support students' elaboration, and the students have opportunities to contribute to extended conversations. Classroom discourse is characterized by the consistent use of feedback/probes that encourage deeper/more meaningful exploration of ideas.

Appendix A, Table 9: Student Peer Feedback

1 - Beginning	2 - Developing	3 - Progressing	4 - Extending
☐ The teacher asks students to assess a peer's work and provide feedback on a trivial task, such as a spelling test, a math facts worksheet, or astate capitals quiz. The task provides no, or almost no, opportunities to comment on the qualityof the work. Rather, the assessment is focused on completeness or accuracy.	☐ The teacher asks students to assess a peer's work on a simple task with few opportunities to provide quality feedback.	☐ The teacher asks students to assess a peer's work and provide feedback on an appropriately complex task to improve thequality of the work.	☐ The teacher asks students to assess a peers' work and provide feedback on a high quality,complex task to improve the quality of the work.
	☐ Few students take the peer feedback task seriously or engage with it meaningfully.	☐ Many students take the peer feedbacktask seriously and engage with it meaningfully.	☐ Most or all students take the peer feedback task seriously and engage with it meaningfully.
	☐ The peer feedback task lacks structureand does not support students. Most students struggle to complete the task and cannot provide feedback that supports learning.	☐ The peer feedback task is structured insuch a way that some students are able to complete the task and provide feedback that supports learning, but the structure may not be adequate for most students.	☐ The peer feedback task is structured in sucha way that most or all students are able to complete the task and provide feedback that supports learning.
	☐ The feedback is of low quality, or no time is provided for students to apply what they learned from the feedback.	☐ Some students receive adequate feedback of adequate quality while othersreceive low-quality feedback, or limited time is provided for students to apply what they learned from the feedback.	☐ All students receive feedback of adequate quality, and sufficient time is provided for students to apply what they learned from the feedback.

Appendix A, Table 10: Student Self-Assessment

1 - Beginning	2 - Developing	3 - Progressing	4 - Extending
☐ The teacher asks students to assess their own learning on a trivial task, such as checking their own work on a spelling test. The task provides no, or almost no, opportunities to commenton the quality of the workor think metacognitively; rather it is focused on completeness or accuracy.	☐ The teacher asks students to assess their own learning on a simple task that only somewhat supports metacognitive thinking.	☐ The teacher asks students to assess their own learning on an appropriate task that is sufficiently complex to support metacognitive thinking focused on improving the quality of their work.	☐ The teacher asks students to assess their own learning on a high-quality, complex taskthat provides many opportunities to support metacognitive thinking focused on improvingthe quality of their work.
	☐ Most students do not take the self-feedback task seriously, or they donot perceive value in the task.	☐ Many students take the self-assessmenttask seriously and engage with it *meaningfully*.	☐ Most or all students take the self-assessmenttask seriously and engage with it *meaningfully*.
	☐ The self-assessment task lacks structureand does not support students (e.g., students do not understand the task or thetask has not been modeled for students). Most students struggle to complete an honest self-assessment.	☐ The self-assessment task is structured ina way that supports some students in completing an honest self-assessment, butthe support may not be adequate for most students.	☐ The self-assessment task is structured in away that supports most or all students in completing an honest self-assessment.
	☐ The output of the self-assessment process does not provide students with evidence that will help them identify ways to improve their work or ways to set goals for further action as appropriate, or the self-assessment may not provide evidence to the teacher aboutstudents' perceptions of their learning in a way that can be used to direct next instructional steps.	☐ The self-assessment process provides students with evidence to help them identify ways to improve their work or setgoals for further action; but students' goals may be vague or not likely to contribute to improvement, or the processmay not provide evidence to the teacher about student perceptions of their learningor be used to direct next instructional steps.	☐ The output of the self-assessment process provides students with evidence by helping them identify ways to improve their work or toset goals for further action as appropriate, or the self-assessment does provide evidence to the teacher about student perceptions of their learning in a way that can be used to direct the next instructional steps.

Appendix B

Appendix B, Table 1: Assessment for Learning BOY Teacher Survey Results

	Usually		Often		Sometimes		About Half the Time		Rarely	
Survey Item	***n***	**%**	***n***	**%**	***n***	**%**	***n***	**%**	***n***	**%**
I post daily lesson-sized learning goals that tell students what they are going to learn in a lesson.	71	61.7	23	20.0	5	4.3	8	7.0	8	7.0
I make meaningful references to my learning goal multiple times throughout the lesson.	29	25.2	42	36.5	17	14.8	19	16.5	8	7.0
I post success criteria that tells students what successful learning will look like.	40	34.8	31	27.0	16	13.9	12	10.4	16	13.9
I provide wait time throughout the lesson.	47	40.9	44	38.3	7	6.1	16	13.9	1	0.9
I collect multiple sources of evidence to identify patterns of understanding and misunderstanding.	29	25.2	45	39.1	15	13.0	17	14.8	9	7.8
When delivering a lesson that students do not understand, I adjust my instruction in the moment.	46	40.0	44	38.3	9	7.8	11	9.6	5	4.3
I model how to use mistakes, failure, and struggle as opportunities for growth.	52	45.2	36	31.3	9	7.8	15	13.0	3	2.6
I provide ongoing feedback to students regarding their use of classroom norms.	39	33.9	46	40.0	10	8.7	16	13.9	4	3.5
I provide frequent opportunities for students to set goals, reflect and monitor their own progress.	22	19.1	27	23.5	21	18.3	19	16.5	26	22.6
I teach students how to give and receive feedback.	14	12.2	19	16.5	27	23.5	20	17.4	35	30.4

Source: Assessment for Learning survey results file dated 06/10/21.
Note: Percentages may not add to 100 due to rounding.

Appendix B, Table 2: Assessment for Learning MOY Teacher Survey Results

	Usually		Often		Sometimes		About Half the Time		Rarely	
Survey Item	***n***	**%**	***n***	**%**	***n***	**%**	***n***	**%**	***n***	**%**
I post daily lesson-sized learning goals that tell students what they are going to learn in a lesson.	40	65.6	15	24.6	2	3.3	3	4.9	1	1.6
I make meaningful references to my learning goal multiple times throughout the lesson.	16	26.2	21	34.4	10	16.4	12	19.7	2	3.3
I post success criteria that tells students what successful learning will look like.	30	49.2	19	31.1	6	9.8	3	4.9	3	4.9
I provide wait time throughout the lesson.	20	32.8	30	49.2	1	1.6	9	14.8	1	1.6
I collect multiple sources of evidence to identify patterns of understanding and misunderstanding.	17	27.9	24	39.3	8	13.1	11	18.0	1	1.6
When delivering a lesson that students do not understand, I adjust my instruction in the moment.	24	39.3	26	42.6	4	6.6	6	9.8	1	1.6
I model how to use mistakes, failure, and struggle as opportunities for growth.	32	52.5	19	31.1	4	6.6	4	6.6	2	3.3
I provide ongoing feedback to students regarding their use of classroom norms.	18	29.5	23	37.7	12	19.7	6	9.8	2	3.3
I provide frequent opportunities for students to set goals, reflect and monitor their own progress.	12	19.7	21	34.4	13	21.3	8	13.1	7	11.5
I teach students how to give and receive feedback.	10	16.4	13	21.3	8	13.1	9	14.8	21	34.4

Source: Assessment for Learning survey results file dated 06/10/21.
Note: Percentages may not add to 100 due to rounding.

Appendix B, Table 3: Assessment for Learning EOY Teacher Survey Results

	Usually		Often		Sometimes		About Half the Time		Rarely	
Survey Item	***n***	**%**	***n***	**%**	***n***	**%**	***n***	**%**	***n***	**%**
I post daily lesson-sized learning goals that tell students what they are going to learn in a lesson.	7	58.3	3	25.0	0	0.0	1	8.3	1	8.3
I make meaningful references to my learning goal multiple times throughout the lesson.	5	41.7	4	33.3	1	8.3	1	8.3	1	8.3
I post success criteria that tells students what successful learning will look like.	8	66.7	2	16.7	1	8.3	0	0.0	0	0.0
I provide wait time throughout the lesson.	6	50.0	3	25.0	2	16.7	1	8.3	0	0.0
I collect multiple sources of evidence to identify patterns of understanding and misunderstanding.	3	25.0	5	41.7	2	16.7	2	16.7	0	0.0
When delivering a lesson that students do not understand, I adjust my instruction in the moment.	4	33.3	6	50.0	2	16.7	0	0.0	0	0.0
I model how to use mistakes, failure, and struggle as opportunities for growth.	7	58.3	4	33.3	0	0.0	1	8.3	0	0.0
I provide ongoing feedback to students regarding their use of classroom norms.	5	41.7	3	25.0	1	8.3	3	25.0	0	0.0
I provide frequent opportunities for students to set goals, reflect and monitor their own progress.	4	33.3	5	41.7	2	16.7	0	0.0	1	8.3
I teach students how to give and receive feedback.	3	25.0	2	16.7	2	16.7	2	16.7	3	25.0

Source: Assessment for Learning survey results file dated 06/10/21.
Note: Percentages may not add to 100 due to rounding.