AIR®
AMERICAN INSTITUTES FOR RESEARCH®

Technology-Mediated Formative Assessment: A Study of Educators' Formative Assessment Practices

SEPTEMBER 2018

Peggy Clements | Whitney Cade
Ann-Marie Faria | Natalie Tucker-Bradway

MAKING RESEARCH **RELEVANT**

AMERICAN INSTITUTES FOR RESEARCH | AIR.ORG

Technology-Mediated Formative Assessment: A Study of Educators' Formative Assessment Practices

SEPTEMBER 2018

Peggy Clements | Whitney Cade
Ann-Marie Faria | Natalie Tucker-Bradway

1000 Thomas Jefferson Street NW
Washington, DC 20007-3835
202.403.5000

www.air.org

4910_07/18

Contents

Tables

Figures

Why This Study?

Educators' interest in using formative assessment to improve student learning increased dramatically after Black and Wiliam's (1998) research review suggested that formative assessment had a large, positive impact on student learning outcomes. More recent research reviews continue to find that students in classrooms where teachers use formative assessment learn more than in classrooms that do not (Kingston & Nash, 2011; Klute, Apthorp, Harlacher, & Reale, 2017). However, not all studies find a positive impact, and the size of the impact may not be as large as Black and Wiliam suggested.

Formative assessment is a process that encompasses more than just assessing students. Even though definitions of formative assessment vary in the specific components included, they all typically describe a multistep process that includes establishing clear learning goals, gathering information about student learning, and using this information to provide feedback and instruction to support student learning (Council of Chief State School Officers, 2008; Heritage, 2010).[1] Many definitions also suggest a relatively short timeline for these cycles, ranging from as short as one that occurs within a single lesson to one that continues over multiple units of instruction (Klute et al., 2017).

Information gathering about student learning—the assessment—may take multiple forms and occur at different points in a lesson or unit (Ciofalo & Wylie, 2008; Heritage, 2007; Herman, Osmundson, Ayala, Schneider, & Timms, 2006; Klute et al., 2017). An assessment may consist of a quiz with multiple questions, a single-item assessment, or even a teacher's observation of students. Questions may have closed- or open-ended responses. In terms of timing, teachers may collect data at the beginning, middle, or end of a lesson; at multiple times during a single lesson; or periodically throughout a series of lessons.

Likewise, how teachers modify their instruction may vary depending on what they learn from the formative assessment data they collect (Heritage, 2007; Herman et al., 2006; Yan & Cheng, 2015). Under some circumstances, teachers may shift the instruction they provide to a class as a whole, for example, changing the pace of instruction or reteaching key concepts to the entire class. Under other circumstances, teachers may choose to provide differentiated instruction to individuals or small groups of students.

Clearly, formative assessment is a complex and ambitious practice. As such, it is critical that school-level supports are in place and barriers to formative assessment are minimized (Faria

[1] Many formative assessment definitions also include having students actively engaged in gathering, interpreting, and acting on information to improve learning, which was not a focus of this study.

et al., 2012). School administrators' leadership and the presence of concrete supports for engaging in formative assessment are important factors for supporting teachers' formative assessment use (Datnow, Park, & Kennedy-Lewis, 2013; Enright & Witham, 2008; Marsh, Pane, & Hamilton, 2006; Murname, Sharkey, & Boudett, 2005). Concrete supports can include clearly stated goals for using formative assessment, as well as providing teachers with the resources and professional development they need to develop and engage in these practices. A strong professional climate in which educators are committed to using data to inform their instruction and identify the best approaches to meeting student needs also is important (Faria et al., 2012; Marsh, 2012; Schildkamp, Poortman, Luyten, & Ebbeler, 2017). Conversely, issues such as the lack of time to use formative assessment in the classroom, the lack of time to collaborate with colleagues, and inflexible curriculum materials can serve as barriers to using formative assessment (Herman, Yamashiro, Lefkowitz, & Trusela, 2008; Ingram, Louis, & Schroeder, 2004).

The process of gathering, processing, and interpreting the information collected about students' learning to inform instruction is challenging for many teachers (Ofsted, 2007; Yan & Cheng, 2015). To facilitate this process, educational technology companies have developed tools to support teachers as they collect, review, and decide how to revise their instruction based on data collected as part of their formative assessment practice. Using technology to facilitate this process builds on recent research studies that indicate the use of educational technology tools to facilitate educators' data collection and review can positively affect students' academic outcomes (Faria et al., 2017; Feng, Roschelle, Heffernan, Fairman, & Murphy, 2014; Roschelle, Feng, Murphy, & Mason, 2016).

One example of an educational technology tool designed specifically to support teachers' use of formative assessment is *Formative*, a Web-based platform and application.[2] Using *Formative*, teachers can create, administer, and grade formative assessments. Teachers can create assessments using a variety of question and response types, ranging from multiple choice to hand-drawn answers. Teachers can then use *Formative* to administer assessments at any point during their instruction and review the data collected either in real time to make in-the-moment shifts in instruction or afterward as they plan instructional next steps. *Formative's* leadership is committed to developing a tool that supports teachers' use of formative assessment with their students. To this end, *Formative's* leadership partnered with the American Institutes for Research (AIR) to conduct a study of their users' experiences using formative assessment as well as their experiences using the *Formative* platform.

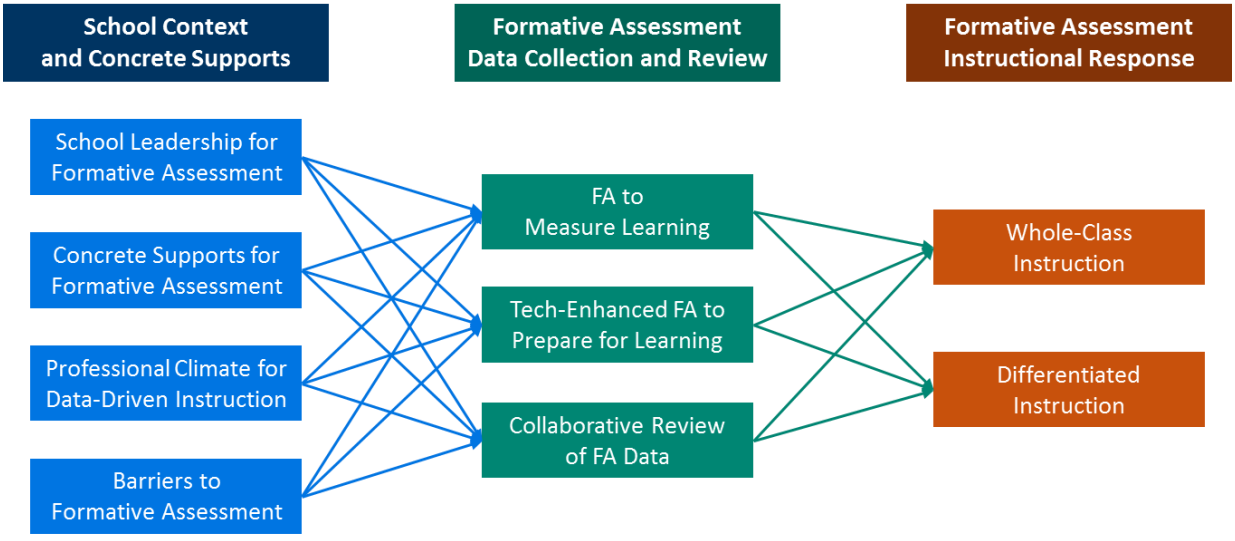
[2] See https://goformative.com for more information.

What the Study Examined

To inform the development of this study, the AIR study team and *Formative's* leadership began by developing a logic model to represent the school-level context and formative assessment practices on which the study would focus. To do this, AIR and *Formative's* leadership adapted a logic model that Faria et al. (2012) developed for an earlier study of data-driven instruction.

As shown in Figure 1, the key factors for the study fell into three categories. The factors in the left column represent school context and concrete supports, and include school leadership for formative assessment, concrete supports for formative assessment, and the professional climate for data-driven instruction. The fourth variable, barriers to formative assessment, represents barriers that may exist at the school level. The variables in the center column and the right column represent two aspects of teachers' formative assessment practices. The factors in the center column measured components of teachers' formative assessment data collection and review. Two scales tapped data collection objectives: formative assessment to measure learning and technology-enhanced formative assessment to prepare for learning. The third measured the extent to which teachers collaborated with colleagues to review formative assessment data. The right column includes scales measuring teachers' formative assessment instructional activities: making changes to whole-class instruction and providing differentiated instruction.

Figure 1. Formative Assessment Logic Model



Note. FA = formative assessment.

To understand how educators use formative assessment, AIR designed and conducted a descriptive and correlational study to provide the developers of *Formative* with information about their users' experiences engaging in formative assessment teaching practices in their schools and their experiences using the *Formative* platform. The study addressed the following research questions:

- How do teachers use *Formative* in terms of frequency?
- Do teachers who are high users and low users of *Formative* differ in terms of their perceptions of the school-level context for engaging in formative assessment practices or their engagement in formative teaching practices?
- What are *Formative* users' perceptions of the common and unique barriers and facilitators to using *Formative*?

How Did the Study Answer the Research Questions?

The AIR study team used multiple data sources and data analysis techniques to answer the research questions.

Data Sources

Three sources provided data for the study: *Formative* system data, individual school characteristics, and an online survey.

Formative System Data

The AIR study team used system data provided by *Formative* to describe the frequency with which *Formative* users logged into the system and identify the sample of teachers who were invited to participate in the survey study.

Data on School Characteristics

A second source of data for the study came from the U.S. Department of Education's National Center for Education Statistics (NCES), which maintains data on all public and private schools in the country. The data include information about each school's setting (e.g., whether it is in an urban, suburban, town, or rural location), student demographic characteristics, and grade level (e.g., elementary, middle, or high school). The AIR study team used these data to describe the school characteristics of all *Formative* users who are teachers, as well as the school characteristics of the sample of teacher users who participated in the survey study.

Online Teacher Survey

The third source of data was an online survey that AIR developed and administered to a sample of *Formative* users who were elementary, middle, or high school teachers in the United States. The survey asked teachers about their experiences using formative assessment and their experiences using the *Formative* platform. To learn about teachers' experiences using formative assessment, the survey included questions to measure the key study factors included in the logic model presented in Figure 1. Table 1 lists the key study factors and sample survey items used to measure each factor.

Table 1. Key Study Factors to Measure Teachers' Experiences Using Formative Assessment and Sample Items for Each Factor

Study factor	Sample items
School context and concrete supports	
School leadership for formative assessment	• The school leadership at my school facilitates conversations about using the formative assessment process. • The school leadership team at my school encourages teachers to make decisions based on formative assessment data.
Concrete supports for formative assessment	• My school provides adequate professional development to give me the skills I need to implement formative assessment. • My school has clear goals for teachers on the use of formative assessment to improve student outcome. • My school's policies enable teachers to address student needs at our campus using formative assessment practice.
Professional climate for data-driven instruction	• Educators in this school use student data to inform changes in instruction. • Educators in this school regularly examine student data. • Educators in this school work hard to match students with interventions that will meet their individual needs.
Barriers to formative assessment	• Curriculum materials are too inflexible. • There is not enough time to engage in formative assessment practices in my classes. • The curriculum doesn't provide materials for formative assessment.
Formative assessment data collection and review	
Formative assessment to measure learning	• To check for concepts and big ideas • To check for progress on standards • At the end of a unit or lesson • At planned checkpoints during a unit or lesson

Study factor	Sample items
Technology-enhanced formative assessment to prepare for learning	• Use digital applications designed for formative assessment. • Check for student understanding before a unit or lesson. • Use formative assessment to check for prior knowledge related to the lesson • Use formative assessment to check for motivation and engagement
Collaborative review of formative assessment data	• Review formative assessment data with teaching teams. • Review formative assessment data in collaboration with campus administrators. • Review formative assessment data informally with other teachers.
Formative Assessment Instructional Practices	
Changes in whole-class instruction	• Change the pace of instruction. • Change the sequence of instruction. • Identify and correct gaps in the curriculum for all students. • Review or reteach key concepts for the entire class.
Differentiated instruction	• Tailor instruction to individual student needs. • Create or adjust same-level achievement groupings of students. • Reteach key concepts to some, but not all students. • Develop or revise individualized education plans.

To measure school context and concrete supports for formative assessment, the survey asked items to determine the extent to which the following factors were present in teachers' schools:

- School leadership for formative assessment
- Concrete supports for formative assessment
- A professional climate for data-driven instruction
- Barriers to engaging in formative assessment

To measure teachers' formative assessment data collection and review practices, the survey included items to measure the extent to which teachers did the following:

- Collected formative assessment data collection to measure student learning
- Engaged in other formative assessment practices, such as using technology tools to collect formative assessment data or collecting data to prepare for instruction
- Collaborated with other teachers and administrators to review formative assessment data

To measure how teachers adapted their instruction in response to the formative assessment data, the survey included items to measure the extent to which teachers did the following:

- Adapted their whole-class instruction
- Provided differentiated instruction to individuals or small groups of students

Teachers responded to all survey items by selecting response options that ranged from 1 (*strongly disagree*) to 4 (*strongly agree*) or from 1 (*not at all*) to 4 (*to a great extent*).

The survey included two other categories of questions. One set of questions asked teachers about their experiences using the *Formative* platform. These questions asked about the types of training teachers received, their perceptions of the tool, and the extent to which they experienced technical difficulties when using *Formative*. The other set of questions asked teachers about their background, such as the grade levels and academic subject(s) they teach and their positions at their schools (e.g., regular education teacher or special education teacher).

Study Participants

The study examined data from two samples of *Formative* users. All teachers in both samples had registered for a *Formative* user account any time between when the company began offering user accounts and before November 15, 2017. In addition, all teachers had logged into the system at least once since August 1, 2017, which the study team established as the beginning of the 2017–18 school year. This ensured that all survey participants were current *Formative* users.

To answer the first research question about the frequency with which *Formative* users log into the system, the study examined data from all *Formative* users who the study team determined were most likely to be K–12 teachers; that is, they indicated a school affiliation and education-related title when registering for their Formative user account (N = 16,933 teachers).

To answer the second and third research questions, the study targeted two types of *Formative* users. One consisted of 250 teachers randomly selected from *Formative* users whose log-in rate after signing up for an account was in the highest quartile. This group of high users had logged into the system between 81 and 852 times in the 1.5 years on average since creating their *Formative* user accounts. The second consisted of 250 teachers randomly selected from *Formative* users whose log-in rate was in the lowest quartile. These low users had logged into *Formative* between two and five times in the .8 years on average since creating their user accounts. Comparing high users and low users allowed the AIR study team to examine whether

any differences existed in the two groups' perceptions of their school contexts for engaging in formative assessment practices, as well as differences in the extent to which they engaged in formative assessment teaching practices. Of the 500 teachers invited to complete the survey, 248 responded to the survey; 140 (56%) were high users, and 108 (44%) were low users.

Appendix B provides additional information about the two groups of study participants: all *Formative* teacher users and the sample of teachers invited to participate in the survey.

Data Analysis

The AIR study team used a variety of data analysis techniques to address the research questions. These included descriptive analyses, exploratory factor analyses and Rasch analyses to develop the scales, ordinary least squares regression analyses to answer the research questions comparing the formative assessment experiences of high users and low users, and structural equation modeling (SEM). Appendix B contains detailed information about how the AIR study team analyzed the data to address each research question.

What Did the Study Find?

This section discusses the study's key findings for each research question.

The Average *Formative* User Logged Into the Platform 72 Times

On average, *Formative* users logged into the platform 72 times since creating a *Formative* user account. The number of logins ranged from one (representing users who did not log into the system again after registering) to more than 1,000 logins. The distribution of user logins was positively skewed, indicating that a larger percentage of users logged in fewer times and a smaller percentage logged in many times. Figure B1 in Appendix B shows the distribution of teachers' total number of logins since establishing their *Formative* user accounts.[3]

High Users and Low Users of *Formative* Did Not Differ in Their Reports of Their School-Level Context for Formative Assessment or in Their Formative Assessment Teaching Practices

Given the wide range in the frequency with which *Formative* users logged into the platform, the study next compared the formative assessment experiences of teachers who had used the

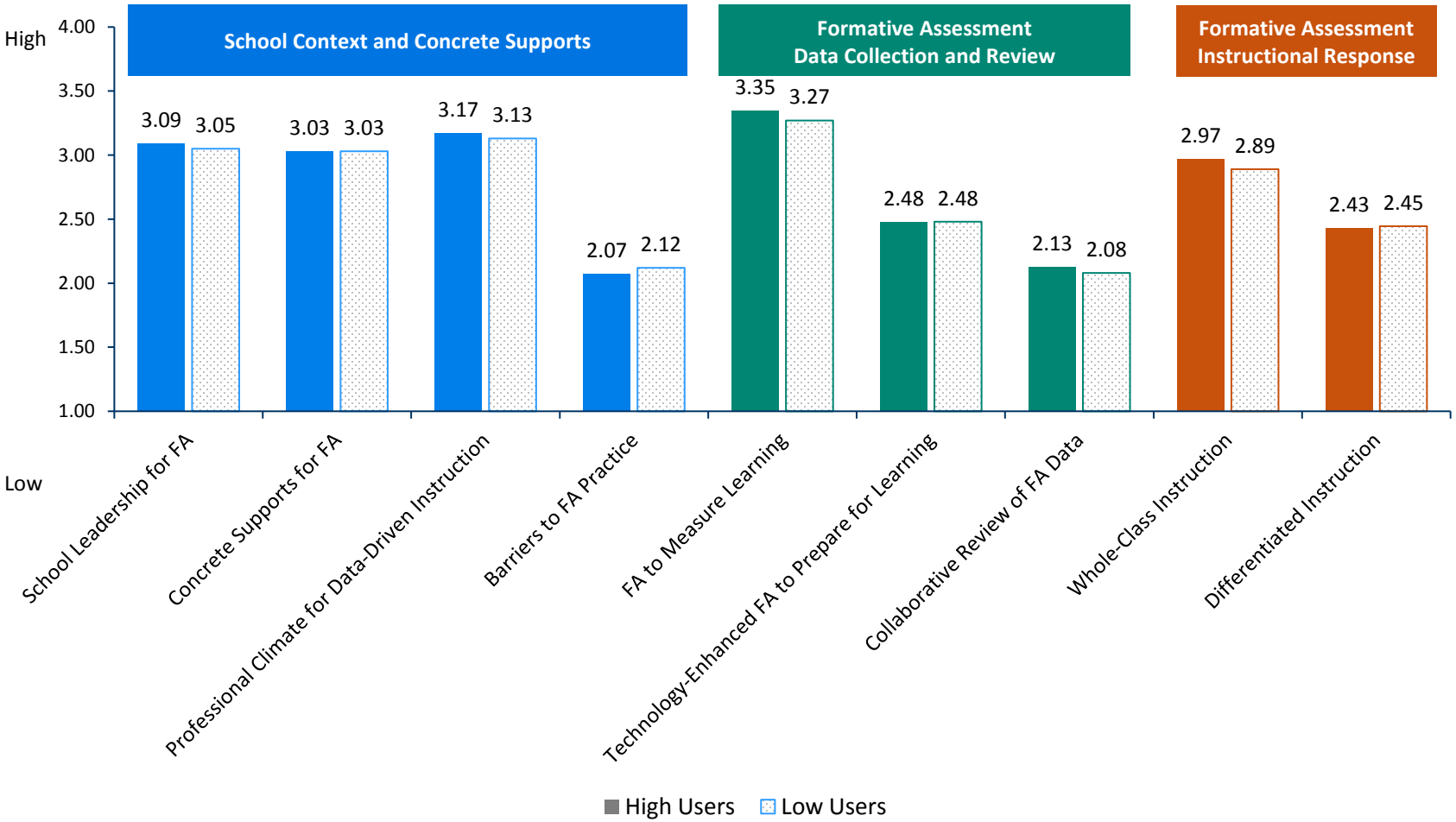
[3] This distribution also is evident in the number of logins for the high users and low users discussed previously when describing the study sample. The lowest quartile represented *Formative* users who had all logged into the system relatively few times (between two and five times total), whereas the highest quartile represented *Formative* users who had a much larger range of logins (between 81 and 852 times total).

Formative platform the most frequently (high users) with teachers who had used it the least frequently (low users). No statistically significant differences were evident between high users and low users on any of the measures of teachers' school context for formative assessment or the extent to which they engaged in formative assessment practices, suggesting that they were more similar to than different from each other in formative assessment, see Figure 2.[4]

- **School Context and Concrete Supports for Formative Assessment.** No statistically significant differences were evident between high users and low users in their reports about the supports for and barriers to engaging in formative assessment at their schools. Both high users and low users agreed, on average, that school leadership and school support for formative assessment were present in their schools, and a professional climate existed for data-driven instruction. In addition, both high users and low users reported encountering barriers to using formative assessment to a minor extent on average.
- **Data Collection and Review for Formative Assessment.** No statistically significant differences were evident between the responses of high users and low users on any of the measures related to formative assessment data collection and review. On average, both groups of teachers reported using formative assessment to track student learning to a moderate extent but reported using formative assessment to assess other academic outcomes to a lesser extent. In addition, teachers reported reviewing formative assessment data in collaboration with their peers to a minor extent on average.
- **Formative Assessment Instructional Practices.** Finally, no statistically significant differences were evident between the responses of high users and low users on either of the measures of formative assessment instructional practices. Both groups of teachers responded that they modified their whole-class instruction to a moderate extent and provided differentiated instruction between a minor and a moderate extent.

[4] The regression analyses for this report used Rasch scale scores as the outcome measures. Analyzing Rasch scale scores is preferable to using scores created by averaging of teachers' responses when the response options represent ordered categories, as was the case for this study. However, Rasch scales can be difficult to interpret as they are not expressed in the same scale teachers used to answer the survey questions. To facilitate the readers' understanding of the results, the figures presented in this report use scale scores created by averaging teachers' responses to across the items.

Figure 2. No Statistically Significant Differences Were Evident Between High Users and Low Users on the Key Study Factors Measuring Teachers' Experiences Using Formative Assessment

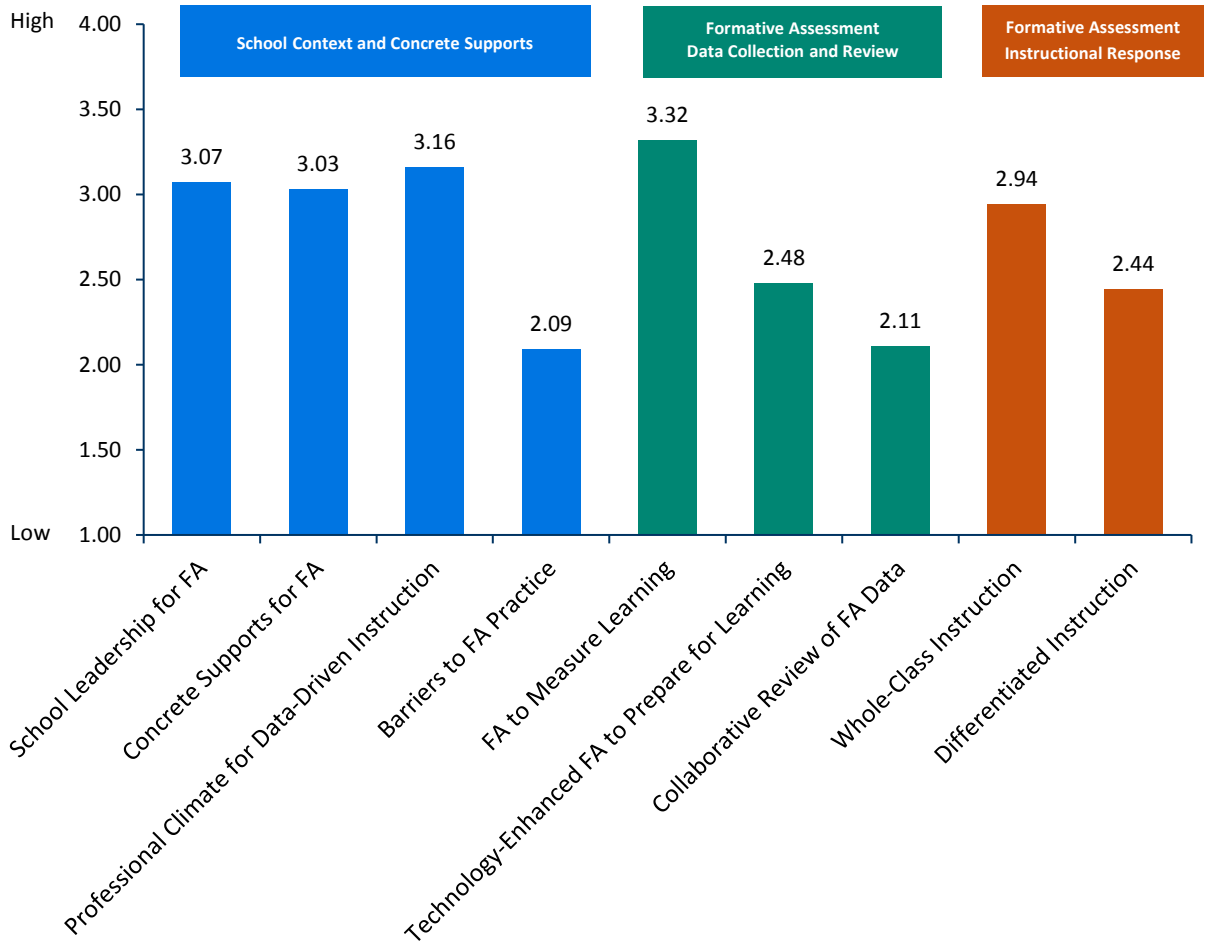


Note. FA = formative assessment.

Given that no statistically significant differences were evident between high users and low users on any of the key factors of interest for the study, the AIR study team examined how all teachers in the study responded to the survey. The goal of these analyses was to better understand the pattern of teachers' responses across the key measures of their formative assessment experiences. Figure 3 suggests the following about teachers' experiences engaging in formative assessment practices:

- Regarding their perceptions of their school contexts, teachers reported relatively high school-level support for engaging in formative assessment practices and relatively low barriers to engaging in formative assessment practices.
- Teachers' responses indicate that they collect information to measure student and progress relatively more frequently than they collect data for other purposes—such as prior to a lesson or to check for prior knowledge—or collaborate with colleagues to review formative assessment data.
- When it comes to revising their instruction in response to the information they have collected about their students, teachers reported revising the instruction they provide to their class as a whole more often than providing differentiated instruction for individuals or small groups of students.

Figure 3. Teachers Reported Lower Levels of Four Key Study Factors: Barriers to Formative Assessment Practice, Technology-Enhanced Formative Assessment to Prepare for Learning, Collaborative Review of Formative Assessment Data, and Differentiated Instruction



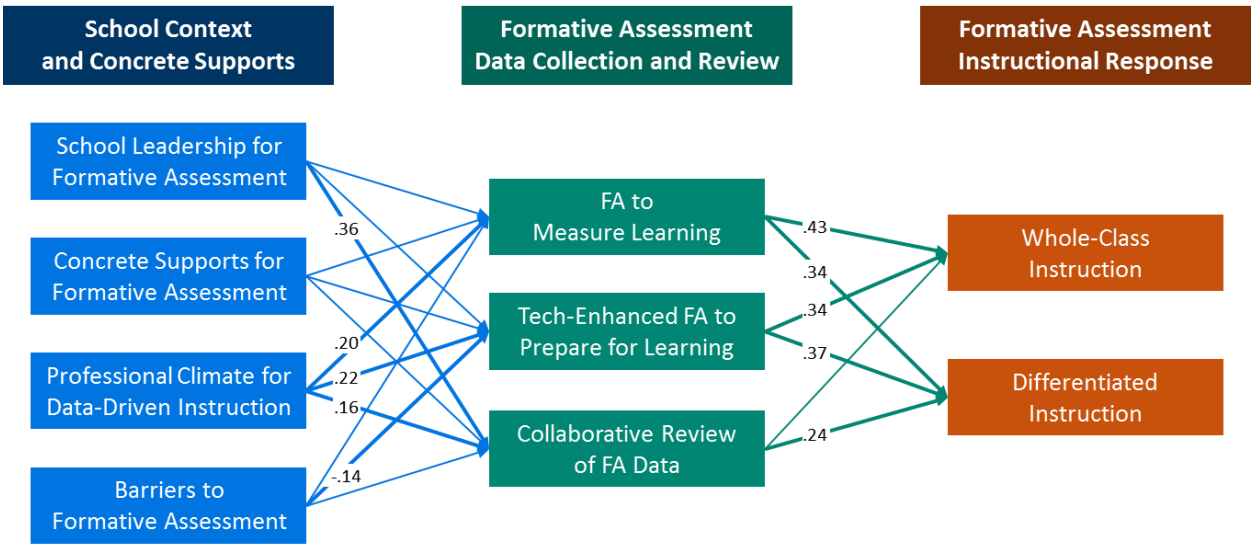
Note. FA = formative assessment. Source: Authors' calculations based on survey data.

An Analysis of All Key Study Variables Confirmed Many of the Relationships Hypothesized in the Logic Model

We next examined the relationships among all the key study factors using SEM. The goal of the analysis was to determine whether empirical support existed for the formative assessment logic model that was developed by the AIR study team and *Formative's* leadership to guide the study. The model examined whether teachers' perceptions of school-level supports for and barriers to formative assessment would predict their engagement in data collection and review practices and whether data collection and review would, in turn, predict teachers' instructional responses.

The estimated structural equation model is presented in Figure 4, with the bold lines representing relationships between variables that were statistically significant.[5] The numbers reported next to each bold line are the standardized regression coefficients representing the strength of the relationship between the two measures. Standardized regression coefficients represent the relationship between two variables in standard deviation units. A standard deviation is the average distance in either direction between all teachers' scores on a scale and the average scale score. One can interpret the standardized coefficients as follows: The estimated standardized relationship between school-level leadership for formative assessment and collaborative review of formative assessment data is .36. This means that, on average, teachers with a school leadership score one standard deviation higher than average have a collaborative review score that is .36 standard deviation units higher than teachers with an average school leadership score.

Figure 4. Structural Equation Model Showing Relationships Among All Key Survey Measures



The overall pattern of predicted relationships was confirmed, although not all relationships in the logic model were statistically significant. The structural equation model indicated the following relationships between school context and teachers' formative assessment data collection and review:

[5] The model also included estimates of the correlations among all the variables in each column. However, the figure does not include lines representing those correlations. Except for the correlation between school leadership for formative assessment and barriers to formative assessment practice, all correlations were statistically significant.

- Of the 12 relationships between school context and concrete supports, five were statistically significant.
- Experiencing a professional climate for data-driven instruction was the most consistent predictor of teachers' reports of data review and collaboration, predicting formative assessment to measure learning, technology-enhanced assessment to prepare for learning, and collaborative review of formative assessment data.
- School leadership for formative assessment predicted teachers' collaborative review of formative assessment data.
- Teachers' reports of experiencing barriers to formative assessment were negatively associated with their use of technology-enhanced formative assessment to prepare for learning but not the other two measures of data collection and review.
- Surprisingly, teachers' reports of concrete supports for formative assessment did not predict any of the data review and collaboration measures.

Five of the six associations between teachers' formative assessment data collection and review practices and the measures of instructional change hypothesized in the logic model were positive and statistically significant. They included the following:

- Teachers' reports of engaging in both types of formative assessment data collection was positively associated with both types of instructional response for formative assessment: whole-class instruction and differentiated instruction.
- The extent to which teachers reported collaboratively reviewing formative assessment data was positively associated with their providing differentiated instruction but not modifying whole-class instruction.

Teachers' Experiences Using *Formative*

The AIR study team also examined how educators interacted with the *Formative* platform, using survey questions about factors that may facilitate or serve as barriers to using *Formative*. To understand some of the factors that may facilitate teachers' use of *Formative*, the survey asked about the training teachers received and their perceptions of the training, as well as teachers' attitudes about the *Formative* platform. To understand possible barriers, the survey asked teachers whether they experienced any technical difficulties when using the *Formative* platform. In this section, we report some of the key findings from these questions. Table B3 in Appendix B reports teachers' responses to all questions about the *Formative* platform.

Teachers answered a series of questions related to training they received to use *Formative* and their perceptions of the training. Forty-seven percent of the teachers who responded to the survey never received training to use *Formative*, 41% received informal training from a colleague, 16% participated in formal training from a colleague, and 5% participated in formal training lead by *Formative* staff.[6] Among teachers who had participated in any type of training, whether formal or informal, 73% agreed or strongly agreed that the training was valuable, and 68% agreed or strongly agreed that the training answered their questions. At the same time, 81% of the teachers indicated that they were interested in learning new or advanced ways of using *Formative,* and 75% indicated they would like to have more training.

Overall, teachers had mostly positive perceptions of *Formative*. Most teachers who responded to the survey believed using *Formative* fit easily into their weekly schedules (59% agreed to a moderate or great extent that it did), was a valuable tool (70%), and improved student learning (61%). Forty percent believed using *Formative* improved student behavior (40%).

Teachers' survey responses suggest that typically they did not encounter technical difficulties when using *Formative*. For example, 85% of the teachers reported that the platform froze either not at all or to a minor extent. When asked whether the *Formative* platform was slow to load, 85% of the teachers reported either not at all or to a minor extent.

Implications of the Study Results

This section first discusses the implications of the study results comparing high-use and low-use teachers, and then discusses the implications of the results of the structural equation model that examined the logic model.

Study Results Comparing High Users and Low Users

One of the study's main objectives was to better understand possible differences between teachers who are frequent and infrequent *Formative* users—the high-use and low-use groups that were the focus of the survey study. However, the high-use and low-use groups appear to be more similar than different in their experiences using formative assessment. No significant differences were found in their reports of their schools' contexts for using formative assessment or in their formative assessment practices.

[6] Teachers could select more than one option, so the sum of the percentages is greater than 100%.

In designing the study, the AIR study team hypothesized that high users and low users would have different experiences using formative assessment overall. At least two reasons explain why the study did not detect differences in the two groups. One is related to the sampling strategy used for the study. In focusing on teachers with *Formative* user accounts, the AIR study team anticipated that the frequency of *Formative* use could serve as a proxy for the extent to which teachers used formative assessment overall. However, it is possible that even though their level of *Formative* use varied, teachers with the initiative to seek out *Formative* and register for a user account are more similar than different. A second possible explanation for the similarity in high-use and low-use teachers' survey responses is that the survey questions did not capture actual differences that may exist in their school contexts or formative assessment practices. Detecting differences may require other data collection methods, such as classroom observation.

Although the comparison of high-use and low-use teachers did not provide insights about teachers' formative assessment use, looking across all teachers' responses provides both promising information and suggestions about where there may be opportunities for improvement. It is encouraging that teachers who have sought out *Formative* reported relatively high school-level support for and lower barriers to engaging in formative assessment practices. It also is promising that teachers reported collecting formative assessment data to measure students' learning to a moderate extent on average. However, teachers' responses on questions about engaging in other formative assessment practices raise questions. For example, teachers may need additional resources or support if they want to use technology-enhanced formative assessment tools or collect information to prepare for learning. Similarly, it is important to understand why teachers reported collaborating with colleagues to review data infrequently. On the one hand, they may prefer to review and make decisions about formative assessment on their own. On the other, it may indicate that teachers need additional support or resources to engage with their colleagues to carry out this work.

Regarding how teachers modify their instruction in response to formative assessment data, the survey results suggest that teachers are better able to adjust their whole-class instruction than they are at providing differentiated instruction. This is understandable, given that differentiating instruction may be more demanding than changing instruction for the whole class. Although formative assessment data may give teachers the information needed to know where students need support, teachers may need more training or additional supports to know how to translate that into personalized instruction.

Evidence for the Study's Logic Model

The results of the structural equation model demonstrated that the correlations among key study factors support the theory of action in general but with some important exceptions. Teachers' perceptions that their school provided a professional climate for data-driven instruction consistently predicted the extent to which they engaged in the formative assessment data collection and review practices measured by the study. In contrast, teachers' perceptions of school-level concrete supports for formative assessment did not predict any of the data collection and review measures, and their reports of school leadership predicted only one measure: collaborating with colleagues to review formative assessment data.

There are several possible explanations for the results. It is possible that the survey questions did not capture the type of information about school contexts that predict teachers' formative assessment data collection and review. Alternately, schools may need to provide more concrete supports for teachers to collaborate to review formative assessment practices. These supports could include creating professional learning communities and providing shared time for data review and reflection. However, an alternate explanation is that how teachers are using *Formative* may not depend on the same school structures that are typically necessary for other types of data use, such as district benchmark data (Faria et al., 2012). *Formative* may lend itself to individual teachers using it with their students because any teacher may download and use *Formative*. One recent study suggested that technology-based innovations may be successful when teachers volunteer versus when they are required to try something new (Petko, Eggers, Cantieni, & Wespi, 2015). To the extent that teachers are choosing to use *Formative* independently, this could minimize their need for the kinds of school-level support necessary for other types of data-driven instruction, such as making good use of district benchmark data.

Study Limitations

When interpreting the results, it is important to keep the study's limitations in mind. First, all results related to teachers' experiences are based on self-reported survey data. Despite being well intentioned, teachers' responses may not accurately reflect their school contexts or instructional activities. Second, as mentioned previously, including only *Formative* users in the study sample may have biased the results such that they do not apply to the larger population of teachers. Third, the study results reflect only the experiences of teachers who responded to the survey. The formative assessment experiences of teachers who chose not to respond may differ from the teachers who did. Fourth, the amount of time since high users and low users had registered for a *Formative* account differed for the two groups. On average, high users had

registered for a *Formative* account 1.5 years ago and low users had registered .8 years ago. Despite this difference, there were no significant differences between the two groups on key study factors, suggesting it did not influence the study's results. Finally, the study used a correlational design, which provides information about how key study factors were related to each other but does not tell us about the causes. For example, the structural equation model demonstrated that teachers' reports of school leadership were correlated with their reports of collaborating with colleagues to review data. This result should not be interpreted to mean that school leadership for formative assessment will lead to higher levels of collaborative data review, but it does indicate that school leadership is associated with teachers' collaborative data review.

Recommendations for Future Research

Moving forward, *Formative* leadership may want to consider implementing other research designs. Given that the goal of having teachers use *Formative* and formative assessment is designed to improve student learning, *Formative* leadership may want to partner with a district planning widespread implementation of the program to conduct additional research. A future study could compare a treatment group of teachers that uses *Formative* and a control group that does not have access to *Formative* on their formative assessment practices and their students' achievement. Ideally, the study would include more objective measures of the school context and teachers' practices, which could provide a better understanding of how teachers use *Formative* and formative assessment in their classrooms. Rather than relying on a self-reported teacher survey data administered once, a future study could collect survey data from multiple sources, classroom observations, or frequent teacher checklists that ask teachers to report on specific instructional activities over a narrow time period.

References

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7–74.

Bond, T. G., & Fox, C. M. (2013). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). London, UK: Psychology Press.

Ciofalo, J. F., & Wylie, E. C. (2006). *Using diagnostic classroom assessment: One question at a time.* New York, NY: Columbia University, Teachers College. Retrieved from http://www.tcrecord.org/PrintContent.asp?ContentID=12285

Council of Chief State School Officers. (2008). *Attributes of effective formative assessment.* Washington, DC: Author.

Datnow, A., Park, V., & Kennedy-Lewis, B. (2013). Affordances and constraints in the context of teacher collaboration for the purpose of data use. *Journal of Educational Administration, 51*(3), 341–362.

Enright, D. M., & Witham, P. (2008). *Examining the existence of and the conditions for teacher commitment and capacity for data use: An exploratory study of MNPS middle school math and language arts teachers* (Capstone Project, submitted to the faculty of Peabody College of Vanderbilt University in partial fulfillment of the requirements for the degree of Doctor of Education in educational leadership and policy). Retrieved from https://discoverarchive.vanderbilt.edu/handle/1803/4004

Faria, A.-M., Heppen, J. B., Li, Y., Stachel, S., Jones, W., Sawyer, K., . . . Palacios, M. (2012). *Charting success: Data use and student achievement in urban schools.* Washington, DC: Council of the Great City Schools.

Faria, A.-M., Sorensen, N., Heppen, J., Bowdon, J., Taylor, S., Eisner, R., & Foster, S. (2017). *Getting students on track for graduation: Impacts of the early warning Intervention and monitoring system after one year* (REL 2017–272). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Midwest. Retrieved from https://ies.ed.gov/ncee/edlabs/regions/midwest/pdf/REL_2017272.pdf

Feng, M., Roschelle, J., Heffernan, N., Fairman, J., & Murphy, R. (2014). Implementation of an intelligent tutoring system for online homework support in an efficacy trial. In S. Trausan-Matu, K. E. Boyer, M. Crosby, & K. Panourgia (Eds.), *Intelligent tutoring systems: 12th international conference proceedings* (Lecture Notes in Computer Science, Book 8474). New York, NY: Springer.

Heritage, M. (2007). Formative assessment: What do teachers need to know and do? *Phi Delta Kappan, 89*(2), 140–145.

Heritage, M. (2010). *Formative assessment: Making it happen in the classroom.* Thousand Oaks, CA: Corwin Press.

Herman, J. L., Osmundson, E., Ayala, C., Schneider, S., & Timms, M. (2006). *The nature and impact of teachers' formative assessment practices* (CSE Technical Report 703). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

Herman, J. L., Yamashiro, K., Lefkowitz, S., & Trusela, L. A. (2008). *Exploring data use and school performance in an urban public school district: Evaluation of Seattle Public Schools' comprehensive value-added system* (CRESST Report 742). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing. Retrieved from https://files.eric.ed.gov/fulltext/ED503302.pdf

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*, 424–453.

Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice, 30*(4), 28–37.

Klute, M., Apthorp, H., Harlacher, J., & Reale, M. (2017). *Formative assessment and elementary school student academic achievement: A review of the evidence* (REL 2017–259). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Central. Retrieved from https://ies.ed.gov/ncee/edlabs/regions/central/pdf/REL_2017259.pdf

Ingram, D., Louis, K. S., & Schroeder, R. G. (2004). Accountability policies and teacher decision making: Barriers to the use of data to improve practice. *Teachers College Record, 106*(6), 1258–1287.

Marsh, J. (2012). Interventions promoting educators' use of data: Research insights and gaps. *Teachers College Record, 114*(11).

Marsh, J. A., Pane, J. F., & Hamilton, L. S. (2006). *Making sense of data-driven decision making in education: Evidence from recent RAND research.* Santa Monica, CA: RAND Corporation.

Murnane, R. J., Sharkey, N. S., & Boudett, K. P. (2005). Using student-assessment results to improve instruction: Lessons from a workshop. *Journal of Education for Students Placed at Risk, 10*(3), 269–280.

Ofsted, G. C. (2007). *Annual report of her majesty's chief inspector, children's services and skills 2006/07*. London, UK: Her Majesty's Stationery Office.

Pane, J. F., Steiner, E. D., Baird, M. D., Hamilton, L. S., & Pane, J. D. (2017). *Informing progress—personalized learning: Teacher and student survey results.* Santa Monica, CA: RAND Corporation. Retrieved from https://www.rand.org/content/dam/rand/pubs/research_reports/RR2000/RR2042/RAND_RR2042z1.pdf

Petko, D., Eggers, N., Cantieni, A., & Wespi, B. (2015). Digital media adoption in schools: Bottom-up, top-down, complementary or optional? *Computers & Education, 84,* 49–61.

Roschelle, J., Feng, M., Murphy, R. F., & Mason, C. A. (2016). Online mathematics homework increases student achievement. *AERA Open, 2*(4). https://doi.org/10.1177/2332858416673968

Schildkamp, K., Poortman, C., Luyten, H., & Ebbeler, J. (2017). Factors promoting and hindering data-based decision making in schools. *School Effectiveness and School Improvement, 28*(2), 242–258.

Snider, R., Fishman, B., Riconscente, M., Tsai, T., & Plass, J. (2015). *Empowering educators: Supporting student progress in the classroom with digital games* (Technical Appendix). Ann Arbor: University of Michigan. Retrieved from http://gamesandlearning.umich.edu/a-games/

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis.* Chicago, IL: Mesa Press.

Yan, Z., & Cheng, E. C. K. (2015). Primary teachers' attitudes, intentions and practices regarding formative assessment. *Teaching and Teacher Education, 45,* 128–136.

Appendix A. Survey Development and Administration

Survey Development

To develop the survey, the AIR study team began by working with *Formative's* leadership to articulate a theoretically and empirically based logic model about the conditions that support teachers' engagement in formative assessment practices, as well as the component instructional practices on which the study would focus. The study team then drew from existing survey instruments to develop a survey that would measure many of the logic model's key components. To capture the breadth of information needed to answer the research questions, the AIR study team adapted items from surveys developed to examine formative assessment as well as related instructional practices, such as data-driven instruction and personalized learning.

The AIR study team used or adapted items from survey instruments from the following studies:

- *Charting Success: Data Use and Student Achievement in Urban Schools* study (Faria et al. 2012)
- *Informing Progress: Insights on Personalized Learning Implementation and Effects* study conducted by the RAND Corporation (Pane, Steiner, Baird, Hamilton, & Pane, 2017)
- *Empowering Educators: Supporting Student Progress in the Classroom With Digital Games* (Snider, Fishman, Riconscente, Tsai, & Plass, 2015)

Teachers responded to all survey items by selecting response options that ranged from 1 (*strongly disagree*) to 4 (*strongly agree*) or from 1 (*not at all*) to 4 (*to a great extent*).

The survey also asked teachers about their background (such as years of experience and education level) and their experiences using the *Formative* platform. To learn about teachers' experiences using *Formative*, the survey included questions about the following:

- The types of training teachers received to use the *Formative* platform
- Teachers' perceptions of the training
- Teachers' perceptions of the *Formative* platform
- The extent to which teachers experienced technical difficulties when using *Formative*

Sample Selection

The AIR study team used online account data that *Formative* collected from all past and present users to identify the two samples used for this study: all *Formative* teacher users and the sample of teacher users invited to complete the survey. The original data file included all 43,253 *Formative* users who agreed to be contacted by a nonprofit research group about participating in a research study. An initial review of all users' information indicated that many registered users were not teachers. This reflects that *Formative* is a free-to-use tool that can serve many purposes for different types of users, not only educators. To ensure that the study focused on only teachers, the AIR study team created the following set of selection criteria to be considered for inclusion in the sample:

- Teach at a school in the United States.
- Have a name and a valid e-mail address on record with *Formative* and must not have marked e-mails from *Formative* as spam or have unsubscribed.
- Indicate that they are a teacher and associated with a school in the United States.
- Not list a nonclassroom role as their primary profession (e.g., rabbi, librarian, or lawyer).
- Have logged into *Formative* at least twice (i.e., at least once to create a profile and one additional time).
- Not have logged on more than 1,000 times (these are rare cases).
- Be associated with only one district.
- Signed up before November 15, 2017, and logged in at least once in the 2017–18 school year (i.e., since August 1, 2017).

This screening process produced the group of 16,993 *Formative* teacher users whose system data the AIR study team examined to address the research question about the frequency with which teachers use *Formative.*

From this group of users, the AIR study team identified a sample of K–12 teachers in the United States to participate in the survey study to answer the research questions about teachers' experiences using formative assessment and their experiences using the *Formative* platform. Drawing from this sample of users, the study defined high users as those whose login frequency was in the highest quartile and low users as those whose login frequency was in the lowest quartile. The AIR study team then randomly sampled and randomly ordered 350 users each from the high user group (n = 4,247) and the low user group (n = 4,296). Beginning at the top of each list, the study team then conducted an Internet search for each user to ensure that he or

she was still actively employed as a classroom teacher in the listed school, verify the teacher's e-mail address, and collect the contact information for the teacher's school. To create the final sample of 500 *Formative* users, this process continued until a list of 250 verified teachers was in each group, all with complete contact information.

The final criterion in the list described previously was an important factor for establishing the sample for the survey study. Because the AIR study team contacted users who had not actively signed up to participate in a study, the study team wanted to ensure that users would remember using *Formative*. The study accomplished this by requiring that users had logged into *Formative* at some point during the school year that the teachers were invited to participate in the survey. In addition, the study team did not sample from very new users. Because one goal of this project was to compare high and lower users, very new users may have had few logins simply because they had little time to use the product.

Survey Administration

The survey was administered during a 6.5-week period in March and April 2018. The AIR study team employed a variety of strategies to encourage participation. Teachers in the study sample received an e-mail invitation with a unique survey link from the AIR study team. The e-mail explained that AIR was contacting them given their interest in formative assessment, which was demonstrated by their having registered for a *Formative* user account. After the initial invitation, the study team used a multifaceted outreach effort to follow-up with teachers. The AIR study team followed up with teachers who had not started the survey or who had not completed the survey by sending one to two reminder e-mails per week. After the survey had been open for 2 weeks, the AIR study team conducted additional follow-up activities by attempting to contact teachers at the school they named when registering for their *Formative* user account. This included making at least one telephone call to teachers who had not started the survey and sending up to two mailed postcards to teachers who had not started the survey or had not completed it. Teachers who completed the survey received a $25 Amazon.com electronic gift card.

Of the 500 teachers invited to complete the survey,

- two hundred thirty-seven answered all survey questions;
- eleven answered some but not all of the survey questions;
- twenty-two did not consent to participate in the study; and
- two hundred thirty did not respond to the survey.

Combining the teachers who answered all or some of the questions, 248 teachers participated in the survey (49.6%). Of those who participated, 140 teachers were from the high-use group (55.5%), and 108 teachers were from the low-use group (45.5%).

Sample Characteristics

This section describes the sample characteristics for the two groups of teachers that the AIR study team examined for this study: the full sample of *Formative* teacher users and the teacher sample for the survey study.

All Formative Teacher Users

The AIR study team used NCES data to describe the full sample of *Formative* teacher users, whose data the study team used to answer the research question addressing how often users log into *Formative*. Using NCES data allowed the study team to provide information about *Formative* users who did not participate in the study, albeit by describing the characteristics of the schools in which they teach rather than the users themselves.[7] The formative user account registration process asks users to select the name of their school where they teach. The AIR study team used this information to link *Formative* users to the NCES data affiliated with their schools and used the NCES data to describe the schools' characteristics.

- *Formative* users teach in middle schools (34%), high schools (30%), elementary schools (28%), and schools that represent other grade structures (8%), such as K–8 or K–12.
- The largest percentage of *Formative* users teach in schools located in areas that the NCES defines as suburban (44%), followed by cities (27%), rural areas (19%), and towns (9%).[8]
- Across all *Formative* users' schools, on average, 47% of the students qualified for the free or reduced-price lunch program.

Formative Teacher Users Who Participated in the Survey Study

As intended given the sampling strategy the AIR study team used to identify teachers to participate in the survey study, respondents taught a wide variety of grades and academic subjects. Of the teachers who responded to the survey, 86% were general education teachers; 7% were special education teachers; and the remaining respondents filled another role at their school, such as administrator, department head or lead teacher, or technology support

[7] In cases where more than one *Formative* teacher user works in a particular school, that school is represented in the analyses multiple times.

[8] The NCES locale codes used for this descriptive analysis indicate a school's location relative to a populous area. More information about the NCES urban-centric locale codes can be found at https://nces.ed.gov/ccd/commonfiles/glossary.asp.

provider. Teachers who responded to the Formative Assessment Survey taught in Grades K–12. Teachers classified as high users most commonly taught Grades 6–8, whereas teacher respondents who were classified as low users most commonly taught Grades 9–12. The academic content areas taught most often by respondents were mathematics (49%), science (39%), reading (36%), and history (31%).[9] Other content areas included foreign or world language, arts, physical education, and career or technical education. The highest education level attained by most respondents was a master's degree (66%), with the remaining teachers having earned a bachelor's degree.[10] The amount of time since teachers had established their *Formative* accounts also varied: 16% had registered for an account more than 2 years ago, 42% registered between 1 and 2 years ago, and 42% registered less than 1 year ago.

Table A1 presents the all results of the descriptive analysis of NCES data for teachers who responded to the survey, the sample of teachers selected to participate in the study, and the population of *Formative* teacher users. In addition to using NCES data to describe *Formative* users, the AIR study team also examined the percentage of teachers in each group who had a premium *Formative* user account. The table presents information for the following groups of *Formative* users:

- High-use teachers who responded to the survey
- Low-use teachers who responded to the survey
- All teachers who responded to the survey (high users and low users combined)
- Teachers sampled to participate in the study but did not respond to the survey
- The full study sample (high users and low users)
- All *Formative* high users
- All *Formative* low users
- The total population of the *Formative* teacher users working as K–12 teachers

[9] The percentages do not sum to 100% because teachers could select more than one subject area.
[10] One respondent reported having earned a doctorate.

Table A1. Average Number of *Formative* Logins, *Formative* Premium Account Status, and School-Level Characteristics of Survey Respondents, the Full Study Sample, and All Formative Teacher Users

Variable	High-use survey respondents (*n* = 140)	Low-use survey respondents (*n* = 108)	All survey respondents (*n* = 248)	All survey nonrespondents (*n* = 252)	All teachers invited to complete the survey (*n* = 500)	All high users (*n* = 4247)	All low users (*n* = 4296)	Total population of *Formative* teacher users (*n* = 16,993)
Premium *Formative* account status								
Percentage with premium account	5%	0%	3%	2%	2%	6%	0%	2%
School-level characteristics								
Pupil:teacher ratio	17:1	16:1	17:1	19:1	18:1	18:1	17:1	17:1
Percentage of students who qualify for free or reduced-price lunch	44%	46%	45%	47%	46%	46%	49%	47%
Mean percentage of male students	52%	51%	52%	51%	52%	51%	51%	51%
Mean percentage of Asian/Pacific Islander students	6%	4%	5%	6%	5%	5%	5%	5%
Mean percentage of Hispanic students	20%	22%	21%	22%	22%	22%	23%	22%
Mean percentage of Black students	9%	12%	10%	14%	12%	12%	15%	13%
Mean percentage of White students	60%	59%	60%	53%	56%	56%	53%	55%
Percentage in a city	20%	20%	20%	28%	24%	26%	27%	27%
Percentage in a rural area	27%	13%	21%	16%	18%	23%	17%	19%
Percentage in a suburban area	36%	55%	44%	49%	47%	41%	46%	44%
Percentage in a town	16%	12%	15%	7%	11%	9%	9%	9%

Variable	High-use survey respondents (*n* = 140)	Low-use survey respondents (*n* = 108)	All survey respondents (*n* = 248)	All survey nonrespondents (*n* = 252)	All teachers invited to complete the survey (*n* = 500)	All high users (*n* = 4247)	All low users (*n* = 4296)	Total population of *Formative* teacher users (*n* = 16,993)
Percentage in elementary schools	25%	21%	23%	24%	24%	25%	28%	28%
Percentage in middle schools	45%	33%	40%	40%	40%	38%	32%	34%
Percentage in high schools	26%	38%	31%	30%	30%	31%	30%	30%
Percentage in other grade structure	4%	7%	6%	6%	6%	6%	9%	8%

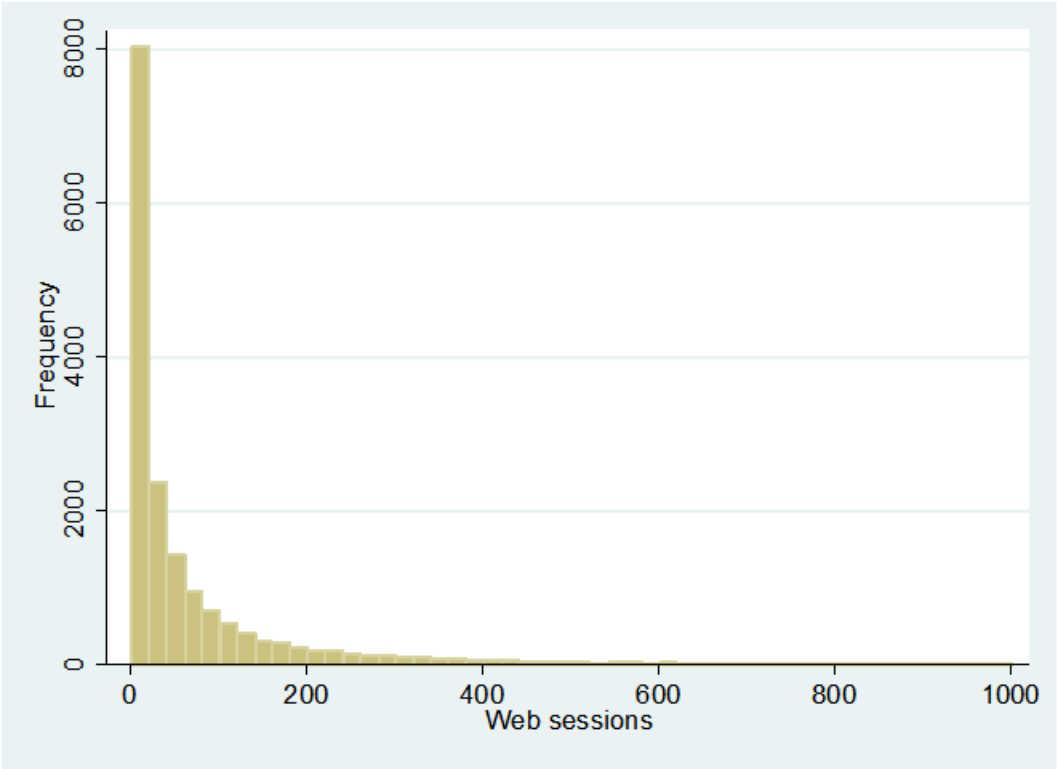
Appendix B. Data Analysis and Results

This appendix provides additional detail about the data analysis methods that the AIR study team used to answer each research question.

Research Question 1: How do teachers use the *Formative* platform in terms of frequency?

To provide information about *Formative* users, the AIR study team conducted descriptive analyses using *Formative* system data to examine the number of times each *Formative* user logged into the system. The number of times teacher users logged into *Formative* ranged from one to more than 1,000 times. Figure B1 presents the distribution of the number of logins, or Web sessions, for the full sample of *Formative* teacher users (*n* = 16,993).

Figure B1. Distribution of the Number of Times All Teacher Users Logged Into Formative Since Creating Their User Accounts



The average number of logins for all *Formative* teacher users since they had registered for their user accounts was 72. Table B1 presents the average number of logins for the total population

of *Formative* teacher users, as well as for the teachers who responded to the survey and the full sample of teachers invited to participate in the survey.

Table B1. Average Number of *Formative* Logins, *Formative* Premium Account Status, and School-Level Characteristics of Survey Respondents, the Full Study Sample, and All Formative Teacher Users

Variable	High-use survey respondents (*n* = 140)	Low-use survey respondents (*n* = 108)	All survey respondents (*n* = 248)	All survey nonrespondents (*n* = 252)	All teachers invited to complete the survey (*n* = 500)	All high users (*n* = 4,247)	All low users (*n* = 4,296)	Total population of *Formative* teacher users (*n* = 16,993)
Frequency of *Formative* use								
Average number of logins	237	3	133	92	225	3	115	72

Research Question 2: Do high users and low users of the *Formative* platform differ in terms of their perceptions of the school-level context for engaging in formative assessment practices or their engagement in formative assessment practices?

To test for differences in the key study measures of teachers' formative assessment experiences between high users and low users of the *Formative* platform, the AIR study team carried out three phases of data analysis. The first phase consisted of exploratory factor analyses (EFA) to discover how the survey items grouped together to measure school context and concrete supports for formative assessment and teachers' formative assessment practices. The EFA identified the groups of survey items to measure each factor. The next data analysis phase used Rasch analyses to assign an overall score on each measure for each teacher who responded to the survey. The Rasch analysis scores provide a more accurate reflection of an individual's responses than a mean score because the Rasch score adjusts for the difficulty of each item (the extent to which teachers agree with an item) rather than treating all items equally. The third step was to conduct ordinary least squares regression analyses using the Rasch score as an outcome to test for differences between the high and low *Formative* users. The following subsections provide additional details about each phase of the analyses.

Exploratory Factor Analysis

EFA identifies underlying patterns in survey items to determine the best fitting construct (or measurement) structure for all items. In other words, EFA identifies sets of items that correlate well with one another but do not correlate as strongly with other items in the survey. To determine the best measurement structure of our survey data, the AIR study team ran the EFA using an orthogonal rotation to minimize the correlations among the resulting factors and estimated multiple models that produced different numbers of constructs overall. The goal was to identify the solution that was aligned with the measurement goals of the survey and retained as many survey items as possible. The study team selected the solution that included 11 factors. In this solution, each item had a factor loading of .40 or greater. Of the 11 factors in this solution, the study team retained eight factors for the Rasch analysis, encompassing 59 of the original 73 items. The study team excluded three factors from further analysis for the following reasons: they were either not interpretable (two factors) or measured a construct that did not address the research questions (one factor). In addition, one of the three factors included multiple items with factor loadings of less than .40.

Rasch Analysis

After the EFA, the AIR study team estimated a partial-credit Rasch model to examine the properties of each construct identified in the EFA (Wright & Masters, 1982). The partial-credit Rasch model assigns an overall, quantitative measure to each teacher based on his or her responses to one or more items within the construct and the difficulty of each item. When items function well together, clear patterns will emerge, from which we can derive the difficulty of each item within a construct as well as the strength of endorsement of the individual on that construct. For example, in a well-functioning construct, we observe survey participants consistently agreeing more strongly to one item and less strongly to another item. With multiple items within the construct, we can then estimate an individual "score" on the construct based on their responses relative to all other survey respondents.

Prior to calculating the overall "score," the AIR study team examined the internal consistency and reliability of each construct to determine if the constructs were functioning as attended. The two measures of reliability, person separation reliability and Cronbach's alpha, measure how consistently survey respondents answered each question. For both measures, reliability ranges from 0 (no reliability) to 1 (perfect reliability). Typically scores above .70 are considered good, and scores above 0.80 even better. As shown in Table B2, most factors had strong reliability, whereas the other factors' reliabilities were adequate for inclusion in the analyses.

Table B2. Factor Reliability Statistics and Item Properties

Factor	Number of items	Person separation reliability	Internal consistency (Cronbach's α)	Range of item fit	Point-correlation range
School context and concrete supports					
School leadership for formative assessment	4	0.80	0.87	0.66–1.25	0.76–0.89
Concrete supports for formative assessment	4	0.80	0.85	0.94–1.16	0.77–0.86
Professional climate for data-driven instruction	5	0.82	0.89	0.57–1.35	0.75–0.87
Barriers to formative assessment	5	0.69	0.76	0.67–1.19	0.65–0.74
Formative assessment data collection and review					
Formative assessment to measure learning	7	0.71	0.82	0.81–1.46	0.47–0.71
Technology-enhanced formative assessment to prepare for learning	6	0.68	0.70	0.80–1.20	0.55–0.69
Collaborative review of formative assessment data	3	0.74	0.73	0.65–1.30	0.77–0.89
Formative assessment instructional practices					
Changes in whole-class instruction	10	0.83	0.87	0.79–1.34	0.59–0.73
Differentiated instruction	15	0.87	0.92	0.57–1.40	0.44–0.74

Next, the AIR study team examined the outfit and point-correlation range for each construct. Outfit measures the extent to which the observed ratings are aligned with expected ratings based on the ratings on all other items (Bond & Fox, 2013). An outfit of 1.00 indicates a perfect fit of an item with the construct, and an outfit between 0.5 and 1.5 is considered good fitting.

Finally, the study team examined the correlation of each item to the overall factor it was used to measure. The analyses suggested that a factor encompassing both school leadership items and school support items could be split into two separate factors, for a total of nine factors. Table B2 lists the final nine factors along with the number of items used to create the scale score, the internal reliabilities (as measured by both the person separation reliability and

Cronbach's alpha), the range of item fit for all items in the scale, and the point-correlations between each item in the scale and the overall scale score. All items had good item fit and were correlated with the overall construct.

Ordinary Least Squares Regression

To examine whether differences existed among high users and low users in their perceptions of their schools' context for formative assessment or in their engagement in formative assessment teaching practices, the AIR study team conducted ordinary least squares regression analyses. For these analyses, the outcome measures were scale scores developed from teachers' survey responses. The analyses statistically controlled for teachers' background characteristics, including years of experience, education level, academic subject(s) taught, and the demographic characteristics of the school in which they taught.[11] To compare the construct scores of high users and low users of the *Formative* platform, the AIR study team ran a series of ordinary least squares regression models using the following equation:

$$ConstructScore_i = \beta_0 + \beta_1 HighUser_i + \sum \beta_n TeacherLevelVariables_i + \sum \beta_n SchoolLevelVariables_i + \varepsilon_0$$

where the *ConstructScore* was the Rasch score for individual *i* (transposed to *mean* = 50 and *standard deviation* = 10) on a given construct; *HighUser* was a binary indicator for whether an individual was a high user of *Formative* (i.e., 1 = high user, 0 = low user); *TeacherLevelVariables* was the set of teacher-level covariates, including years of teaching experience, education (i.e., master's degree or higher), and subject taught; and *SchoolLevelVariables* was the set of school-level covariates, including school type (i.e., elementary, middle, or high school), locale (i.e., urban, rural, suburban), the demographic characteristics of the students attending the school, and the student-to-teacher ratio. The coefficient of primary interest in the model was β_1, which represents the difference between a high user and a low user on the construct of interest after controlling for all other variables in the model. As noted in the report, the regression analyses indicated that no statistically significant differences were evident between high users and low users on any of the constructs.

Structural Equation Modeling

The study used SEM to examine whether the associations among all key study factors provided evidence for the logic model that the AIR study team and *Formative's* leadership developed to

[11] The 248 teachers who participated in the survey taught at 240 schools. Eight schools are represented twice in the data. In six schools, both teachers were high users; in one school both teachers were low users; and in one school one teacher was a high user, and one teacher was a low user.

guide the research. SEM is a general statistical modeling technique that can be used to analyze the relationships among multiple variables simultaneously. It allows researchers to hypothesize direct and indirect relationships among a set of variables and test whether evidence exists that the model represents the relationships among the variables observed in the data. SEM also estimates the strength and statistical significance of the relationships specified by the model.

The AIR study team estimated and compared the fit of two SEM models. The first model, in addition to estimating the relationships among the variables specified in the logic model, also included estimates of the correlations among all the school context and concrete supports measures (those on the left side of the logic model) and between the two formative assessment instructional response measures. Estimating the correlations among all variables on the left and right sides of a model is the default mode for most statistical software used for SEM. The AIR study team then assessed the fit of this model following Hu and Bentler's guidelines (1998), focusing on the standardized root-mean-square residual (SRMR) and the root-mean-square error of approximation (RMSEA), as well as χ^2, and determined that the model as specified did not provide an adequate fit to the data. The study team then estimated a second model that added correlations among the three measures of formative assessment data collection and review, the factors in the center column of the logic model. The fit statistics for this model (SRMR = 0.02, RMSEA < 0.01, χ^2 = 7.73, degrees of freedom = 8, p = .46[12]) indicated that the model provided a good fit of the data.

Research Question 3: What are teachers' perceptions of the common and unique barriers and facilitators to using *Formative*?

The AIR study team examined the frequencies for all survey items that measured teachers' perceptions of the barriers and facilitators to using *Formative*. Table B3 presents the frequencies for each survey item used to answer this research question.

Table B3. Frequencies for Items Measuring Facilitators and Barriers to Using *Formative*

Item	Strongly disagree	Disagree	Agree	Strongly agree
Perceptions of training				
The training prepared me well to use the system.	4%	23%	62%	11%
The amount of time allocated to the training felt right to me.	5%	32%	52%	11%

[12] The nonsignificant χ^2 indicates that the covariances of the estimated model do not differ significantly for the observed covariances among all the variables in the analysis, indicating model fit.

Item	Strongly disagree	Disagree	Agree	Strongly agree
The training did a good job of answering my questions.	2%	20%	68%	10%
The training was useful.	2%	9%	70%	20%
I am interested in learning about new or advanced ways of using *Formative*.	2%	9%	58%	31%
I would like to have more training.	3%	21%	56%	19%

Item	Not at all	To a minor extent	To a moderate extent	To a great extent
Perceptions of *Formative*				
It is easy to fit using *Formative* into my weekly schedule.	8%	33%	31%	28%
I believe that my use of *Formative* improves the behavior of students.	20%	40%	29%	11%
I believe that my use of *Formative* improves the achievement of students.	10%	30%	42%	19%
I am committed to using *Formative* in my classroom.	14%	31%	32%	24%
Compared with other systems or programs I've tried before, *Formative* is a valuable.	7%	23%	38%	32%

Item	Not at all	To a minor extent	To a moderate extent	To a great extent
Barriers to using *Formative*				
If I want to use *Formative*, I have to use my personal time to review the data.	11%	21%	33%	35%
Formative freezes frequently.	43%	42%	11%	4%
Formative automatically logs me off, and I need to log back in frequently.	65%	27%	6%	1%
If I contact technical support, they cannot address my problem.	82%	12%	4%	2%
Formative is slow to load.	46%	39%	8%	7%
I have trouble logging in to *Formative.*	78%	16%	4%	2%
My school Internet connection does not allow me to connect to *Formative.*	82%	14%	3%	1%

Item	Strongly disagree	Disagree	Agree	Strongly agree
There are not enough functioning computers/tablets at my school to use *Formative*.	81%	12%	2%	5%

AIR®

AMERICAN INSTITUTES FOR RESEARCH®

Established in 1946, the American Institutes for Research (AIR) is an independent, nonpartisan, not-for-profit organization that conducts behavioral and social science research on important social issues and delivers technical assistance, both domestically and internationally, in the areas of education, health, and workforce productivity.

MAKING RESEARCH **RELEVANT**

AMERICAN INSTITUTES FOR RESEARCH
1000 Thomas Jefferson Street NW
Washington, DC 20007-3835 | 202.403.5000
www.air.org

LOCATIONS

Domestic: Washington, DC (HQ) | Monterey, Sacramento, and San Mateo, CA | Atlanta, GA | Honolulu, HI | Chicago and Naperville, IL Indianapolis, IN | Metairie, LA | Waltham, MA | Frederick and Rockville, MD | Chapel Hill, NC | New York, NY | Columbus, OH | Cayce, SC Austin, TX | Reston, VA

International: El Salvador | Ethiopia | Haiti | Honduras | Zambia

4910_07/18