

Integrating Educational Technology with Formative Assessment in Grades 6–8 Mathematics Instruction: A Quantitative Synthesis of Achievement Outcomes and Moderating Factors

¹Chisom Gabrielle Ezechukwu

¹ St. Bernadette Parish School, Burien, Washington, USA

Email: gabbyeze333@gmail.com

Orcid ID: <https://orcid.org/0009-0001-3612-8351>

DOI: <https://doi.org/10.63084/cognexus.v1i01.309>

Abstract

This study presents a quantitative synthesis of the existing literature regarding the integration of educational technology (EdTech) tools into formative assessment practices within grades 6-8 mathematics classes. Based upon Vygotsky's concept of the Zone of Proximal Development and the formative assessment structure established by Black and Wiliam (1998, 2009), the study collected outcome data from 12 different studies that included over 28,000 students. The results of these studies show that the use of EdTech tools within mathematics classes results in improved outcomes for those students, with effect sizes ranging from Hedges' $g = 0.09$ to Cohen's $d = 0.63$. Also, moderator analyses revealed that variables such as the duration of the implementation of the technology, the alignment of the technology with the curriculum and the fidelity of the teachers in implementing the technology were all associated with the size of the effect sizes. Also, analyses of the specific types of technology revealed that curriculum-integrated technologies produced the largest effects sizes ($d = 0.50-0.63$; Roschelle & Shechtman, 2013), while intelligent tutoring systems (ITS) produced smaller but still statistically important effect sizes ($g = 0.01-0.09$; Steenbergen-Hu & Cooper, 2013). Finally, analyses of the psychometric properties of digital formative assessment instruments revealed high levels of internal consistency ($\alpha = .91$; Fadhila, 2024) and important gains in student scores following the implementation of these assessments ($t = 8.214$, $p < .001$). So, these findings have important implications for educators, curriculum designers and policymakers who wish to integrate technology into their mathematics classrooms.

Keywords: formative assessment, educational technology, middle school mathematics, intelligent tutoring systems, Grades 6–8, quantitative synthesis

Introduction

Formative assessment is the ongoing process of evaluating student learning to inform instruction. The basic review of formative assessment by Black and Wiliam (1998) revealed that implementing formative assessments in classrooms resulted in important improvements in student achievement. Specifically, their review of over 250 studies found that the implementation of formative assessments in classrooms led to gains in student achievement of between 0.4 and 0.7 standard deviations. Following this basic study, two extra studies published by Black and Wiliam (2009) and Wiliam (1999, 2000) revealed that the implementation of formative assessments in classrooms can be accomplished through five specific strategies: clarifying learning intentions, designing classroom discussions, providing feedback that moves learning forward, activating learners as mutual resources and supporting student ownership of learning.

Despite the strong theory and evidence supporting the use of formative assessments in mathematics education, the implementation of such assessments in middle schools remains inconsistent. The main reasons cited by teachers for the lack of implementation of formative assessments are time constraints, limited ability to diagnose student learning and the availability of assessment tools (Troy & Bulgakov-Cooke, 2013). Also, grades 6-8 are an important period in the development of students' mathematical knowledge, during which they transition from arithmetic to pre-algebra and algebra (Roschelle & Shechtman, 2013). So, identifying mechanisms for implementing formative assessments in these grades is both urgent and necessary for improving the learning outcomes of these students.

Educational technology platforms have introduced new capacities for collecting data-rich, formative assessments of students' knowledge. Intelligent tutoring systems, adaptive learning platforms, simulation platforms, game-based response systems and digital assessment platforms all allow teachers to collect formative assessment data from students in real-time and to provide feedback to those students immediately (Feng et al., 2006; Koedinger et al., 2010). The question, however, is whether the implementation of such technology platforms has led to statistically and practically important gains in the mathematics knowledge of students who are exposed to such technology platforms and under what conditions such gains are most likely to happen.

This research study will investigate this question through a quantitative synthesis of the existing literature published through 2025. Three research questions will be used to structure the study:

1. What is the magnitude of the achievement gains for students who use technology-enhanced formative assessments in their mathematics classes in Grades 6-8?
2. How do the magnitudes of outcomes vary according to the type of technology platform that is employed in the education of students?

3. What moderating variables (e.g., implementation duration, curriculum alignment, learner performance level, teacher fidelity) are associated with the variation in effect size magnitude among the studies?

Review of Literature

Formative Assessment: Theoretical Foundations

Black and Wiliam's (1998) synthesis of the literature on formative assessment revealed that it was a high-yield instructional strategy that differed from summative evaluation both in its purpose and processes. Formative assessment, therefore, was an key component of the learning process. A subsequent publication by Black and Wiliam (2009) revealed three basic learning processes and three learning agents (the teacher, the learner and the peers), resulting in a 3 x 3 matrix of formative assessment strategies that has become the dominant structure for understanding formative assessment. Formative assessments are used in mathematics to identify misconceptions regarding number sense, algebraic reasoning, proportional thinking and geometric understanding (Anderson et al., 2012). The alignment study conducted by Anderson et al. (2012) on the easyCBM Curriculum-Based Measurement system revealed that the middle school mathematics formative assessments had 87.0% direct alignment to the Common Core State Standards (post rater-effect control) and that 99.6% of the items on these assessments addressed one of the standards or the prerequisite skills for that standard. So, the assessments were able to provide the necessary information to accurately diagnose students' knowledge of mathematical concepts.

Wiliam (1999) stated that feedback is the most extensively studied mechanism of formative assessment. For feedback to be effective in the classroom, it must be provided in a timely manner, be specific to the skills that the student is attempting to learn and be actionable by the student. Wiliam (2000) further stated that the most enduring gains from formative assessments happen when students are given agency and ownership of their learning, a concept that is similar to the concept of self-regulated learning.

Educational Technology Platforms for Formative Assessment in Grades 6–8 Mathematics

The earliest attempts to implement formative assessments within mathematics classrooms through technological means originated in efforts to integrate the concepts of intelligent tutoring systems (ITSs) into assessment software. Feng et al. (2006) created the E-ASSISTment system, a web-based application that assessed the knowledge of over 600 middle school students within the 2004-2005 school year, yet also provided those students with tutoring feedback as they took the assessments. This type of system became the template for the development of similar ITS-based formative assessments tools. Koedinger et al. (2010) evaluated the effectiveness of the ASSISTments system among approximately 1,240 seventh grade students from three different schools that used the system, as well as one control school. Students who utilized the system had higher scores on their post-assessment tests

compared to students from the control school and that the higher scores were correlated with the number of times that each student utilized the system. Lastly, Zheng et al. (2019) attempted to use formative assessment data gathered from over 23,000 Grade 6-8 students that used the MATHia ITS system over three different school years. Using several different statistical analysis techniques, the researchers discovered that the formative assessment data gathered by the system could be effectively used to predict the scores that the students would earn on their end-of-year math tests.

The most compelling evidence regarding the impact of technology on mathematics learning within the targeted grade band is found within the SimCalc studies conducted by Roschelle and Shechtman (2013). Three large-scale studies published the results of students' learning gains after implementing SimCalc into their classrooms, with effect sizes ranging from $d = 0.63$ to $d = 0.56$. SimCalc incorporates simulations into the mathematics curriculum along with the teachers who implement the program, showing its status as a maximally-integrated technology with formative assessment. Steenbergen-Hu and Cooper (2013) published a systematic review of 26 studies that evaluated the impact of ITS on the mathematics knowledge of students between kindergarten and high school. The average effect size of those studies ranged from $g = 0.01$ to $g = 0.09$. Also, the authors determined that the effectiveness of ITS programs was highest with implementations that lasted less than one school year and with general student populations relative to those of lower achieving students within those populations.

More recently, Möller (2024) conducted a study on the Bettermarks ITS that involved 2,700 students and analyzed their results from approximately five million problem attempts made before, during and after the COVID-19 pandemic. The results indicated that students' math scores increased significantly during all three time periods, especially those with lower initial math scores. Also, other digital formative assessments have been developed for secondary mathematics. Fadhila (2024) published a study on the validation of a digital formative assessment tool for secondary mathematics, which yielded a validity mean of 4.62, a Cronbach's alpha of .91, a teacher practicality mean of 4.47, a student practicality mean of 4.58 and a paired-sample t-statistic of 8.214 ($p < .001$). These results show the tool's validity and practicality for both teachers and students. Finally, game-based formative response systems are the most recent category of EdTech integration. Rahim et al. (2025) conducted a study on the use of Kahoot as a formative assessment tool for Grade 5 students. The results indicated that the students who used the Kahoot tool scored significantly higher than the control group, with means of 25.52 ($SD = 1.74$) and 14.59 ($SD = 2.17$), respectively and a Cohen's d of 4.92. Although the study only involved Grade 5 students and the effect size is notably large, these results do show that the tool can be effective in increasing students' math scores.

Pham et al. (2024) investigated the use of Microsoft Forms as a platform for formative assessments in seventh-grade mathematics classes with 116 students and 30 teachers who were interviewed in Vietnam. The results indicated that the students had improved their mastery of algebraic concepts and that the

teachers found the platform to be very practical. Also, Brezack et al. (2025) conducted a study with 59 mathematics teachers from grades 6-8 in a platform that provided formative assessments to students and discovered that the teachers' understanding of the self-regulation features of the platform were positively correlated with the engagement of their students.

Implementation Fidelity and Equity Considerations

These studies show that the implementation fidelity of EdTech programs has an effect upon the outcome of those programs. For instance, Troy and Bulgakov-Cooke (2013) discovered that only nine teachers out of the total number of teachers in a given school district had maintained the required data records over a period of two years. Similarly, Gillespie et al. (2022) discovered that the transition of formative assessment tools to online platforms had accelerated the adoption of such tools by teachers; however, the lack of digital access to those tools by some students created an inequity within the classroom that negatively impacted the learning of those students.

Research Gap

Despite the growing body of evidence that has been reviewed above, three gaps remain in the literature. First, no meta-analysis has been conducted that specifically aggregates the effect sizes of studies of Grades 6-8 students. Second, while some meta-analyses have investigated moderators of the effectiveness of educational technologies (e.g., Steenbergen-Hu & Cooper, 2013), none of those studies have investigated the effect of the features of formative assessments that are implemented in those technologies. Finally, while some studies have investigated the predictive validity of educational technologies (Zheng et al., 2019), none of those studies have investigated the effectiveness of those technologies in improving student instruction. So, the current study aims to address each of these three gaps in the literature.

Theoretical Framework

The following two theoretical constructs are used to describe the synthesis of these educational technologies. The first theory is that of Vygotsky's concept of the Zone of Proximal Development (ZPD). This theory shows that students learn best within the range of their abilities when provided with scaffolding tools and techniques. Educational technologies that use the ZPD concept to improve learning include systems that provide feedback to students on their performance relative to the target performance for the lesson, as well as systems that provide scaffolding tools to students based upon their individual learning needs. These tools can be considered as a means of providing scaffolded instruction to students. The second theory is that of Black and Wiliam's (1998, 2009) five-strategy formative assessment model. Each of the technologies described above can be mapped onto this model as follows: ITS systems are mapped onto strategies 1 and 3; simulation platforms are mapped onto strategy 2; and game-based response systems are mapped onto strategies 4 and 5. So, these theories

allow for an evaluation of both the outcomes of these educational technologies as well as the extent to which they map onto established theories of learning.

Method

Research Design

This study employed a quantitative research synthesis design to aggregate, analyze and interpret outcome data from studies that investigated the use of EdTech tools in assessing student knowledge in mathematics classes for students in Grades 6 through 8. A quantitative research synthesis is the most appropriate research methodology for this study as it aims to summarize the numeric outcomes of various studies and to determine the factors that moderate the outcomes of these studies (Steenbergen-Hu & Cooper, 2013). The protocol for this study included a four-stage procedure: searching databases, screening the results for eligibility, extracting data from eligible studies and summarizing the data from those studies with a comparison of the moderators of the outcomes of those studies.

Inclusion and Exclusion Criteria

Studies were included in this review if they met all five of the following criteria: (1) publication or presentation year between 1998 and 2025; (2) the study employed a quantitative or mixed-methods research design that included extractable numeric outcomes; (3) the study's primary content area was mathematics; (4) the study sample included students from Grades 6-8; and (5) the study employed at least one identifiable EdTech tool that served a formative assessment function. Studies were excluded from the review if they employed a qualitative research design only, did not report any measure of student learning or psychometric outcomes or if the study sample consisted of students in post-secondary or primary education levels (Grades K-5) only. Application of these criteria yielded 12 studies for inclusion: 10 primary empirical studies, one K-12 meta-analysis, and one technical psychometric alignment report.

Data Extraction and Analytic Approach

The following data were collected from each of the studies that were to be reviewed: the authors and year of publication of each study, the sample size of students that were analyzed in each study, the target grade levels of the students in each study, the type of EdTech platform or tool that was used in each study, the type of study design that was performed in each study, the primary outcome measure that was used in each study and the effect size or inferential statistic that was calculated for each study. For studies that calculated effect sizes based on measures like Cohen's d or Hedges' g , those values were used in the analysis of each study. For studies that only provided inferential statistics (t , F , p), the direction of the study results was coded for analysis but no attempt was made to analyze the magnitude of those results.

Results

Descriptive Overview of Included Studies

Table 1 presents the descriptive profile of the 12 studies that are eligible for inclusion in this review. These studies together represent a sample of over 28,000 learners from different countries. The individual sample sizes of the studies range from 59 (teacher-level; Brezack et al., 2025) to over 23,000 (Zheng et al., 2019). The study designs included randomized controlled trials, quasi-experiments, longitudinal studies, psychometric validations, mixed-methods studies and one meta-analysis of K-12 studies. The seven types of technology platforms that were identified include intelligent tutoring systems (ASSISTments, MATHia, Bettermarks), simulation-based curricula (SimCalc), game-based response systems (Kahoot), digital assessment tools, formative assessment tools (Microsoft Forms), curriculum-based measurement tools (easyCBM) and district implementations of multiple tools.

Table 1

Quantitative Profile of Studies on Technology-Enhanced Formative Assessment in Grades 6–8 Mathematics (1998–2025)

Study	Year	N	Grade Band	Platform / Tool	Design	Key Quantitative Outcome
Black & Wiliam	1998	>250 studies	All	Formative assessment (general)	Systematic review	ES range: $d = 0.40–0.70$
Feng, Heffernan & Koedinger	2006	>600	6–8	E-ASSISTment (ITS)	Longitudinal assessment study	Concurrent validity established
Anderson, Irvin, Alonzo & Tindal	2012	Item sample	6–8	easyCBM	Psychometric alignment	87.0% standard alignment; 99.6% skill coverage
Koedinger, McLaughlin & Heffernan	2010	1,240	7	ASSISTments (ITS)	Quasi-experimental	Significant post-test advantage ($p < .05$); dose-response confirmed
Steenbergen-Hu & Cooper	2013	26 reports	K–12	ITS (varied)	Meta-analysis	$g = 0.01–0.09$
Roschelle & Shechtman	2013	Large-scale	6–7	SimCalc (curriculum-integrated)	2 RCTs + 1 quasi-experiment	$d = 0.63$; $d = 0.50$; $d = 0.56$

Troy & Bulgakov-Cooke	2013	9 teachers	6–8	Multiple tools	Implementation evaluation	Variable; no reliable state-score impact detected
Zheng, Fancsali, Ritter & Berman	2019	23,046	6–8	MATHia (ITS)	Predictive modeling	Reliable summative score prediction (linear, logistic, RF models)
Gillespie, Winn, Faber & Hunt	2022	Unspecified	7	Online FA tool	Longitudinal case study	Variable implementation; access gaps identified
Möller	2024	~2,700	5–9	Bettermarks (ITS)	Longitudinal evaluation	Significant performance increase; larger gains for low performers ($p < .05$)
Fadhila	2024	Secondary cohort	7–9	Digital FA instrument	R&D validation	Validity $M = 4.62$; $\alpha = .91$; $t(\text{paired}) = 8.214$, $p < .001$
Pham et al.	2024	116 students; 30 teachers	7	Microsoft Forms	Mixed-methods (quant)	Significant improvement in algebraic concept mastery
Rahim, Mohammed & Batool	2025	~60	5	Kahoot (game-based CRS)	RCT (pre-test/post-test)	$M_{\text{exp}} = 25.52$ ($SD = 1.74$) vs. $M_{\text{ctrl}} = 14.59$ ($SD = 2.17$); $d = 4.92$
Brezack et al.	2025	59 teachers	6–8	Formative FA platform	Survey/observational	Positive SRL-engagement association

Note. N = total analytic sample. RCT = randomized controlled trial. ITS = intelligent tutoring system. FA = formative assessment. CRS = classroom response system. SRL = self-regulated learning. RF = random forest. For Black & Wiliam (1998), the row represents a synthesized effect size range from a narrative review, not a primary study. The Rahim et al. (2025) study is at Grade 5; it is included as adjacent evidence and excluded from Grades 6-8-specific ES calculations. "Large-scale" for Roschelle

& Shechtman (2013) shows studies with large, representative samples; exact N was not reported in the reviewed abstract.

5.2 Effect Sizes by Technology Platform Type

Table 2 organizes the effect size data that was extractable from the studies by platform category. Four categories of platforms emerged with estimable effect sizes: (a) ITS, (b) curriculum-integrated simulation platforms, (c) game-based classroom response systems and (d) digital formative assessment instruments. Each of these effect sizes is descriptive only, as the confidence intervals for each could not be calculated due to the heterogeneity of the statistics that were reported in each study.

Table 2

Comparative Summary of Effect Sizes by Educational Technology Platform Type

Platform Category	k (Studies)	Cumulative N	Reported Effect Size(s)	Approximate Mean ES	Grade Coverage	Notes
Intelligent Tutoring Systems (ITS)	4	~27,586	$g = 0.01-0.09$ (Steenbergen-Hu & Cooper, 2013); significant $p < .05$ (Koedinger et al., 2010; Möller, 2024)	$\sim g = 0.05-0.09^a$	K-12; Grades 5-9; Grade 7	Small but statistically reliable; dose-response confirmed
Curriculum-Integrated Simulation	1	Large-scale	$d = 0.63$; $d = 0.50$; $d = 0.56$ (Roschelle & Shechtman, 2013)	$d = 0.56$	Grades 6-7	Largest and most consistent ES in the synthesis
Game-Based Response System	1	~60	$d = 4.92$ (Rahim et al., 2025)	$d = 4.92^b$	Grade 5	Exceptionally large ES; small N ; Grade 5 only; interpret with caution
Digital FA Instruments	3	~116 (primary)	$t = 8.214$, $p < .001$; 87.0% alignment; sig. mastery gains (Fadhila, 2024; Anderson et al., 2012; Pham et al., 2024)	— ^c	Grades 7-9; Grades 6-8	Psychometric quality high ($\alpha = .91$); no uniform d reported

Note. *k* = number of contributing studies per platform category. ^a Mean ITS effect size is estimated from the Steenbergen-Hu and Cooper (2013) meta-analysis range; individual primary ITS studies (Koedinger et al., 2010; Möller, 2024) reported importance but not *d* or *g*. ^b The Rahim et al. (2025) *d* = 4.92 reflects a small-sample RCT design and is not used in cross-platform mean comparisons. ^c Digital FA instrument studies reported *t*-statistics and practicality ratings rather than Cohen's *d*; effect sizes cannot be meaningfully averaged across the category. FA = formative assessment.

5.3 Moderating Variables

Table 3 presents the five moderating variables identified across the synthesis, along with the direction of their association with outcome magnitude and the evidentiary basis for each pattern.

Table 3

Summary of Moderating Variables and Directional Associations with Achievement Outcomes

Moderating Variable	Categories Observed	<i>k</i> Reporting	Direction of Association	Supporting Studies
Technology platform type	ITS vs. curriculum-integrated simulation	5	Curriculum-integrated > ITS	Roschelle & Shechtman, 2013; Steenbergen-Hu & Cooper, 2013
Implementation duration	Short-term (< 1 year) vs. sustained (≥ 1 year)	3	Short-term > long-term for ITS; sustained > short for curriculum-integrated	Steenbergen-Hu & Cooper, 2013; Roschelle & Shechtman, 2013
Learner performance level	Low-performing vs. general population	2	Low-performing students show larger ITS gains	Möller, 2024; Steenbergen-Hu & Cooper, 2013
Curriculum/standard alignment	High alignment (≥ 87%) vs. unvalidated	2	High alignment → stronger predictive and instructional validity	Anderson et al., 2012; Zheng et al., 2019

Teacher implementation fidelity	High and consistent vs. variable/low	2	High fidelity → detectable achievement impact	Troy & Bulgakov-Cooke, 2013; Brezack et al., 2025
---------------------------------	--------------------------------------	---	---	---

Note. k = number of studies providing evidence for the moderating variable. All directional associations were determined through comparisons between the results of different studies; no meta-regression analysis was performed due to the differences in the statistical analyses performed within each study. So, these results should be considered as a means of generating hypotheses regarding the relationship between the moderator and outcome variables.

Discussion

Platform Type as a Primary Differentiator of Effect Magnitude

The first major finding of this review is that there are differences in the effect sizes produced by simulation-based learning platforms that are integrated into the curriculum versus standalone ITS platforms. The study by Roschelle and Shechtman (2013) revealed that three different studies that employed the SimCalc software experienced effect sizes ranging from $d = 0.50$ to $d = 0.63$, which is above the threshold of $d = 0.40$ that Black and Wiliam (1998) have established as the standard for measuring the effectiveness of formative assessments. Also, these studies utilized a teaching method that ensured that teachers could incorporate the formative assessments into their lesson plans, collect the assessment results and adjust their lesson plans according to the students' performance. In contrast, standalone ITS platforms experienced average effect sizes ranging from $g = 0.01$ to $g = 0.09$ (Steenbergen-Hu & Cooper, 2013). The main difference between these two types of learning platforms is that the teacher-mediated response to the formative assessments was incorporated into the curriculum-based simulation platforms but not incorporated into the standalone ITS platforms. However, if the ITS platforms were altered to provide teachers with actionable, actionable reports regarding the performance of their students, then the learning outcomes would improve; for instance, Koedinger et al. (2010) found that 1,240 seventh-grade students who used the ASSISTments platform experienced statistically important improvements in their test scores following the implementation of those teacher-level reports.

Predictive Validity and Early Warning Capacity

Another important contribution of ITS to formative assessment comes from Zheng et al.'s (2019) study of the prediction of summative assessment outcomes using instruction-embedded formative data collected from more than 23,000 learners in grades 6 through 8. The researchers used both linear regression and random forest algorithms to determine that formative data could be used to accurately

predict summative assessments for these students. Also, because this prediction was based upon data that was collected during instructional activities rather than during summative assessments themselves, the study shows that these formative assessments could potentially serve as a means of identifying students who may struggle with summative assessments prior to their administration. However, the accuracy of such predictions is only as accurate as the formative assessments themselves; Anderson et al.'s (2012) study of the easyCBM assessment indicated that 99.6% of the items contained within those assessments related to either the Common Core standards themselves or to concepts that are required to understand those Common Core standards. So, these assessments contain valid measures of student knowledge and understanding, showing their potential use as formative assessments within the ITS platforms.

Psychometric Quality of Digital Formative Instruments

Fadhila's (2024) study of the validation of a digital instrument for formative assessments of secondary school mathematics has provided strong indicators of the instrument's reliability and validity. The Cronbach's alpha value of .91 far exceeds the .80 value typically required of instruments that are to be used in high-stakes formative assessments. Also, the paired-sample t value of 8.214 ($p < .001$) shows that the instrument is responsive to changes in the instruction that is to be taught. Also, the means for both teachers (4.47) and students (4.58) in relation to the instrument's practicality are both near the ceiling of a five-point scale, showing that the instrument is both very valid and easy to implement in classroom settings. Each of these properties of the instrument show that it is a digital formative assessment instrument that is capable of providing the reliable data necessary for the making of consequential instructional decisions.

Implementation Fidelity as an Effect Moderator

Implementation fidelity emerged as one of the most important factors within the research literature. For instance, Troy and Bulgakov-Cooke (2013) reported that the use of a technology for formative assessments led to decreased effectiveness of that technology due to the low level of implementation fidelity by teachers; only nine teachers out of the full district had sufficient data to assess the effectiveness of the technology. Also, Brezack et al. (2025) reported that teachers who were more competent in using the formative assessment technology had higher levels of student engagement in their classrooms. Finally, Black and Wiliam's (2009) argument that formative assessments are the responsibility of teachers, rather than the technology itself, is also reflected within these studies. While the technology reduces the effort required to collect and review data regarding students' knowledge and skills, teachers must still interpret the results of those assessments in order to appropriately adjust their classrooms. As such, professional development of teachers is necessary to enable them to appropriately implement formative assessments in their classrooms.

Differential Effects for Low-Performing Learners

Möller's (2024) finding that lower-performing students exhibited greater gains from the use of ITS systems compared to higher-performing students, which is corroborated by the results of Steenbergen-Hu and Cooper's (2013) meta-analysis, has important implications for equity in the classroom. In large classrooms containing students of various skill levels, students with lower performance levels typically receive less feedback from teachers compared to their higher performing peers. However, the adaptive ITS system provides unlimited, differentiated and immediate feedback to students, showing that these systems can provide an important service to students with lower performance levels. Therefore, strategies can be employed to deploy these systems in classrooms with a high proportion of lower performing students.

Game-Based Assessment

The notably large effect size of 4.92 reported by Rahim et al. (2025) for Kahoot must be considered within the setting of the study's small sample size and grade level (Grade 5). The effect size likely results from the small sample size and the natural floor-ceiling effect of the game-based assessment tool. Nonetheless, the findings are consistent with the theories of game-based assessments and engagement in learning environments. Wiliam (1999) discussed the importance of providing students with timely and engaging feedback during assessments to motivate them to learn. So, while the current study provides some insight into the potential effectiveness of Kahoot, the gap in the existing literature regarding strong, large-sample studies of Grades 6-8 students utilizing game-based assessments is a high-priority area for future research.

Conclusion

This synthesis of research studies shows that the integration of educational technology with formative assessments in mathematics classes for students in grades 6 through 8 results in measurable benefits for those students. The magnitude of these benefits also varies across the different types of technology platforms used in these studies. For instance, studies that used technology platforms that were integrated into the curriculum for those subjects showed the largest and most consistent benefits for students who used those platforms ($d = 0.50-0.63$), whereas studies that used information technology systems (ITS) platforms showed smaller but still statistically important benefits for those students ($g = 0.01-0.09$). Studies also show that these formative assessments predict end-of-year summative assessments (Zheng et al., 2019) and that the digital instruments used in these studies were of high psychometric quality ($\alpha = .91$; Fadhila, 2024). Based on these findings, three recommendations can be made regarding the use of technology in formative assessments in mathematics classes. First, the largest benefits for students will be achieved when technology is integrated into the curriculum for those classes. Second, the professional development of teachers should include training on how to interpret the data from those

assessments and how to respond to that data in their teaching practices. Finally, before any digital formative assessment tool is implemented in a classroom, it should be evaluated according to psychometric standards (e.g., content validity, internal consistency and treatment sensitivity) to ensure that the tool is of high quality.

Despite the findings of these studies, there are several limitations to the research that was conducted. For instance, the studies were all conducted in different ways and reported their results differently. Also, there are no studies that were conducted specifically with students in the 6th through 8th grades, which limits the generalizability of the findings of these studies. Also, there are few studies that examine the impact of these games on specific groups of students, such as those who are English language learners or those from low-income backgrounds. So, future studies should aim to conduct large-scale randomized trials of these types of games in the 6th through 8th grades, investigate the effects of these games over time and examine how these games may have an impact on students of different demographic backgrounds. Overall, though, the evidence that does exist regarding these games shows that they are effective in improving the math skills of students who play them, especially when they are implemented with the right curriculum and teacher support.

References

- Anderson, D., Irvin, P. S., Alonzo, J., & Tindal, G. (2012). *The alignment of the easyCBM middle school mathematics CCSS measures to the Common Core State Standards* (Technical Report No. 1208). University of Oregon, Behavioral Research and Teaching.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2002). *Working inside the black box: Assessment for learning in the classroom*. King's College London, Department of Education and Professional Studies.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31. <https://doi.org/10.1007/s11092-008-9068-5>
- Brezack, N., Yang, S., Grady, J., Lavine, A. L., Feng, M., & Li, L. (2025). Teacher perspectives on student self-regulated learning in a math formative assessment platform. In *Proceedings of the 2025 ACM LAK Conference*. ACM. <https://doi.org/10.1145/3698205.3733949>
- Fadhila, N. (2024). Development of a digital formative assessment instrument for mathematics learning in secondary education. <https://doi.org/10.62872/0vxb0y82>

Feng, M., Heffernan, N. T., & Koedinger, K. R. (2006). Addressing the testing challenge with a web-based e-assessment system that tutors as it assesses. In *Proceedings of the 15th International Conference on World Wide Web* (pp. 307–316). ACM. <https://doi.org/10.1145/1135777.1135825>

Gillespie, J., Winn, K., Faber, M., & Hunt, J. H. (2022). Implementation of a mathematics formative assessment online tool before and during remote learning. *Lecture Notes in Computer Science*, 168–173. https://doi.org/10.1007/978-3-031-11647-6_29

Koedinger, K. R., McLaughlin, E. A., & Heffernan, N. T. (2010). A quasi-experimental evaluation of an on-line formative assessment and tutoring system. *Journal of Educational Computing Research*, 43(4), 489–510. <https://doi.org/10.2190/EC.43.4.D>

Möller, K. (2024). *Performance increases in mathematics within an intelligent tutoring system during COVID-19 related school closures: A large-scale longitudinal evaluation*. Martin Luther University Halle-Wittenberg. <https://doi.org/10.25673/115378>

Pham, T., Duong, H. T. T., Nguyen, A. H., Nguyen, H. H., Tran, Y. H., Le, A. V., Ngo, L., & Pham, H. S. T. (2024). Applying Microsoft Forms software in formative assessment in teaching mathematical concepts for 7th graders in Vietnam. *Tạp chí Giáo dục*, 203–217. <https://doi.org/10.52296/vje.2024.400>

Rahim, M., Mohammed, L. A., & Batool, S. (2025). The effect of Kahoot as a gamification-based assessment tool on primary students' academic achievement in mathematics. *International Journal of Technology in Education*, 8(4), 977–997. <https://doi.org/10.46328/ijte.1220>

Roschelle, J., & Shechtman, N. (2013). SimCalc at scale: Three studies examine the integration of technology, curriculum, and professional development for advancing middle school mathematics. In S. J. Hegedus & J. Roschelle (Eds.), *The SimCalc vision and contributions* (pp. 125–143). Springer. https://doi.org/10.1007/978-94-007-5696-0_8

Steenbergen-Hu, S., & Cooper, H. (2013). A meta-analysis of the effectiveness of intelligent tutoring systems on K–12 students' mathematical learning. *Journal of Educational Psychology*, 105(4), 970–987. <https://doi.org/10.1037/a0032447>

Troy, T., & Bulgakov-Cooke, D. (2013). *Formative assessment with technology 2011–12: Second year of implementation* (D&A Report No. 13.05). Chicago Public Schools, Department of Research, Evaluation, and Assessment.

William, D. (1999). Formative assessment in mathematics part 2: Feedback. *Equals: Mathematics and Special Educational Needs*, 5(3), 8–11.

William, D. (2000). Formative assessment in mathematics part 3: The learner's role. *Equals: Mathematics and Special Educational Needs*, 6(1), 19–22.



Zheng, G., Fancsali, S., Ritter, S., & Berman, S. R. (2019). Using instruction-embedded formative assessment to predict state summative test scores and achievement levels in mathematics. *Journal of Learning Analytics*, 6(2), 153–174. <https://doi.org/10.18608/JLA.2019.62.11>