

The Relationship between Formative Assessment and Summative Assessment in Primary Grade Mathematics Classrooms*

Tuba Gezer^a, Chuang Wang^{*b}, Andrew Polly^c,
Christie Martin^d, David Pugalee^e, Richard Lambert^f

Received : 21 February 2021
Revised : 24 May 2021
Accepted : 15 June 2021
DOI : 10.26822/iejee.2021.220

* This project was supported by the North Carolina Department of Public Instruction's Mathematics and Science Partnership grant program.

^aTuba Gezer, University of North Carolina at Charlotte, Charlotte, USA.
E-mail: tgezer@uncc.edu
ORCID: <https://orcid.org/0000-0001-9819-4146>

^bCorresponding Author: Chuang Wang, University of Macau, Taipa, Macau, China.
E-mail: wangc@um.edu.mo
ORCID: <https://orcid.org/0000-0002-3810-9277>

^cAndrew Polly, University of North Carolina at Charlotte, Charlotte, USA.
E-mail: drew.polly@uncc.edu
ORCID: <https://orcid.org/0000-0003-2370-4409>

^dChristie Martin, University of South Carolina, Columbia, USA.
E-mail: martinc1@mailbox.sc.edu
ORCID: <https://orcid.org/0000-0001-7896-6882>

^eDavid Pugalee, University of North Carolina at Charlotte, Charlotte, USA.
E-mail: dkpugalee@uncc.edu
ORCID: <https://orcid.org/0000-0002-3356-1600>

^fRichard Lambert, University of North Carolina at Charlotte, Charlotte, USA.
E-mail: rglander@uncc.edu
ORCID: <https://orcid.org/0000-0002-5791-962X>

Abstract

This study used hierarchical linear modeling to examine the relationship between an internet-based mathematics formative assessment and data from a mathematics summative assessment for primary grade learners (ages 5-7). Results showed a positive relationship between formative assessment data related to the concepts of counting and decomposing numbers and summative data. This relationship was more robust in classrooms where students demonstrated lower average performance on the formative assessment data. The results suggest that formative assessment can be more beneficial to encourage low achieving students in primary-grade mathematics classrooms. Therefore, we recommend teachers to use formative assessment practices more frequently in low achieved primary grade classrooms. The formative assessment process includes the cycle of data collection, data analysis, planning future instruction, and examining the impact of that instruction through cycling back to data collection. This study contributes to the field by providing more empirical data about the relationship between formative and summative assessment with primary grades' learners.

Keywords:

Elementary Education, Formative Assessment, Hierarchical Linear Modeling, Mathematics Education, Summative Assessment, Number Sense

Introduction

Teaching mathematics effectively continues to be a challenge (Chapman, 2012). Thames and Ball (2010) stated that specific skills and knowledge are required to engage in the complex process of teaching mathematics. In this sense, required knowledge refers to pedagogical and content knowledge; however, employing this knowledge to create a classroom that facilitates effective instruction also requires knowing the students. One of the benefits of assessment is learning about student progress. Connor



Copyright ©
www.iejee.com
ISSN: 1307-9298

© 2021 Published by KURA Education & Publishing.
This is an open access article under the CC BY-NC-ND license. (<https://creativecommons.org/licenses/by/4.0/>)

et al. (2018) articulated that teaching mathematics in small flexible learning groups using assessment data to individualize instructions was associated with a significant increase in students' mathematics achievement. Classroom assessment techniques, which embed assessment within the instructional process, help teachers to understand better students' understanding and misconceptions (Veldhuis & Heuvel-Panhuizen, 2019).

In classrooms, formative assessment is used to gather data about the student learning process. Formative assessment and the effective use of data about students' performance may also lead to long-term gains on mathematics assessments (Polly et al., 2017; Black & William, 1998; NCTM, 2014). Formative assessment is important because it is situated within the learning process, and the purpose of formative assessment is not for summative measurement. Heritage (2007) concurred that the use of formative assessment instruments could yield information about students' learning that teachers can use directly to inform future instructions. Also, she emphasized that formative assessment should be situated within a learning paradigm and not as another test within a measurement paradigm.

While formative assessment uses student data to improve teaching and learning, summative assessment uses data to evaluate the learning outcomes (American Educational Research Association, American Psychological Association, & the National Council on Measurement in Education [AERA, APA & NCME], 2014). Formative assessment is usually informal, ongoing, and during instruction, whereas summative assessment is tended to be formal, cumulative, and after instruction (Dixson & Worrell, 2016). The purpose of the study is to examine the relationship between formative assessment and a state summative mathematics assessment. The researchers further examined the relationship between formative and summative assessment based on the average classroom achievement in primary grades (ages 5-7).

Literature Review

Formative assessment is considered an assessment for learning because the target audience of formative assessment is teachers and students, while summative assessment is considered as an assessment of learning due to giving updated information to stakeholders about teacher and student performances (Burke, 2010; Heritage, 2013). The distinguishing characteristics of summative assessment are classifying students' performances and being administered at the end of a unit or semester, whereas the goal of formative assessment is to determine students' strengths and

weaknesses so that teachers and students could decide the beneficial activities to reach educational goals (Cizek, 2010).

Formative assessment is the most frequent assessment with the smallest scope, whereas summative assessment is the least frequent assessment with the most substantial scope, including teachers and school districts. Interim assessments are between formative assessment and summative assessment in terms of frequency and scope (Perie et al., 2009). Teachers commonly use interim assessment to identify the weakness of content or students for planning purposes (Riggan & Olah, 2011). On the other hand, an advantage of formative assessment is active class participation (Randel et al., 2016).

Perie et al. (2009) posited that the label of formative assessment was used to describe interim assessments, and Shepard (2008) suggested considering the purpose of the instrument when labeling formative assessment. Shepard's focus is on the interactions between the teacher and students when labeling an assessment, and she suggests that an interim assessment may produce the interactions that would be considered formative and situated in the learning paradigm. This framework means that when formative assessments are analyzed over time, they can offer the same benchmark insights as interim assessments.

The Relationship between Formative Assessment and Summative Assessment

Formative assessment and summative assessment are commonly used in educational settings to give information regarding student achievement (Hattie, 2003). However, the implications of formative assessment and summative assessments are different. While formative assessment is strongly tied to local curriculum and administered according to students' needs (Shepard et al., 2018), summative assessment uses data to assess students' knowledge (American Educational Research Association, American Psychological Association, & the National Council on Measurement in Education [AERA, APA & NCME], 2014). The concern of covering all the curricular objectives to prepare students for end-of-year summative assessment may influence teachers' formative assessment practices (Box et al., 2015; Govender, 2019).

In the seminal Cognitively Guided Instruction (CGI) project, teachers participated in professional development to learn about students' mathematical thinking, listen and notice students' thinking, and adjust their future work based on the data (Fennema et al., 1996). Thus, the goal of CGI is to help teachers to understand their students' mathematical thinking in order to make instructional decisions based on

students' thinking (Fennema et al., 1996), which indicated that teachers use the data as a formative assessment tool. The researchers investigated 21 teachers' instruction and beliefs about students' thinking for four years and concluded that the changes in teachers' instruction improved students' mathematic achievement (Fenema et al., 1996). In addition, as students' mathematic achievement increase, teachers were encouraged to use CGI strategies to understand students' mathematical thinking (Carpenter & Franke, 2004). Similarly, Stewart (2016) analyzed the predictive validity of formative assessment in 3rd-grade mathematics classrooms. They concluded that formative assessment, which is unit-based curriculum assessments, predicted students' performance on state-wide academic readiness assessment.

Formative Assessment and Feedback

McMillan (2010) articulated that formative assessment is shaped by educational goals and contextual factors such as classroom environment, grade level, or student ability. Besides, Heritage (2007) stated that by design, formative assessment could give feedback on multiple levels, such as providing feedback to teachers and students regarding students' learning processes. Also, Heritage mentioned that feedback could have a strong influence on student motivation and self-efficacy. Faber et al. (2016) examined the effect of digital formative assessment tools on student motivation and success using a randomized experimental design. The feedback feature of formative assessment tools contributed to students' mathematics achievement. The researchers further concluded that teachers benefited from the feedback feature of digital formative assessment more than their students. Similarly, Atjonen (2014) found that teachers emphasize the importance of using multiple assessment methods and interactive techniques to provide constructive feedback. Some researchers noted that delayed feedback might benefit high achieving students, especially with complicated tasks, whereas low achieving students need immediate feedback (Mason & Bruning, 2001). These studies (Atjonen, 2014; Faber et al., 2016; Mason & Bruning, 2001) highlighted that feedback is a crucial aspect of formative assessment for teachers and students.

Crossouard and Pryor (2012) stated that formative assessment includes examining small tasks and controlling observable behaviors. The researchers explained the formative assessment process as describing learning goals, setting assessment criteria, and providing feedback. Crossouard and Pryor suggested using open-ended questions in order to support students' high-level thinking. Similarly, Clark (2010) highlighted the quality of the interaction

between teacher and student and concluded that feedback is a part of formative assessment as long as it leads students to critical thinking to reach learning goals. When process-oriented feedback is provided multiple times during the unit, formative assessment may influence student perception of the usefulness of the assessment (Rakoczy et al., 2019). Additionally, formative assessment can increase students' learning and student motivation (Faber et al., 2016).

Use of Technology in Formative Assessment

Pachler et al. (2010) defined "formative e-assessment" as processes that involve technology that produce data about students' understanding connected to objectives. This data enables both the teacher and the learner to take action to increase learning and improve teaching. This definition shows that any technology with the right conditions can be used for formative assessment. The benefit of using online classroom response systems in formative assessment presents information to teachers and students right after and during instructional practices (Irving, 2015). Considering the time limit during classroom activities, Ramsey and Duffy (2016) concluded that two strengths of using technology in the formative assessment process were saving time for interactive learning activities and facilitating individualized learning. Elmahdi et al. (2018) examined the usage of a tech-tool, Plickers, in formative assessment and concluded that using technology-based tools could improve the quality of formative assessment and students' learning outcomes. Besides examining four aspects of formative assessment, feedback, discussions, personalized options, and game-based learning, interactive whiteboards demonstrated a significant positive correlation between formative assessment activities and math achievement (Chen et al., 2020).

Numerous online formative assessment tools are available to support teachers' planning implementation of modified instruction. While some instruments function as an adaptive learning program by assessing students' outcomes as well as instructional activities, others provide formative assessment data to teachers so that teachers can evaluate students' learning outcomes and individualized instructional strategies. AMC Anywhere is an online-based formative assessment tool that allows teachers to collect data through diagnostic interviews regarding students' learning of mathematics concepts (Richardson, 2012). AMC Anywhere is used in one-on-one settings, and during these sessions, students use manipulatives such as counters and snap cubes to demonstrate number concepts.

The goal of AMC Anywhere is to provide information to teachers about students' conceptual understanding

of number sense so that teachers could use the information to modify instructional activities (Richardson, 2012). The AMC Anywhere assessment tool provides continuous feedback during the assessment process so that teachers would not waste time on too easy or difficult assessments (Martin & Polly, 2015). The tool assists the teachers in combining assessment data, generate individual or class reports according to assessment data, and collaborating with their colleagues (Polly et al., 2016). To sum up, the benefits of the AMC Anywhere tool are saving time on assessment practices, improving the quality of collaboration and feedback, and serving as formative assessment in mathematics classrooms (Martin & Polly, 2015).

Teachers who used AMC Anywhere reported challenges of finding the time and using formative assessment data (Martin, Polly et al., 2016). Despite these difficulties, teachers reported modifying instructional activities based on their students' mathematics achievement (Polly et al., 2016). Further, a positive relationship was reported between the frequency of applied formative assessment activities and students' mathematics achievement, and frequent formative assessment is also to make students know about their own processes (Polly et al., 2018).

Number Sense

Both national and state standards in the primary grades require teachers to focus on developing their students' understanding of the numbers, their relations, and number systems, symbolizing numbers (Common Core State Standards for Mathematics, 2010; National Council of Teachers of Mathematics, 2000; <state blinded> Department of Education, 2017). Number sense is defined as "a child's fluidity and flexibility with numbers, the sense of what numbers mean and an ability to perform mental mathematics and to look at the world and make comparisons" (Gersten & Chard, 1999, p. 18). The primary focuses of early childhood education are counting, addition, subtraction, and understanding of place value (Richardson, 2012). However, children sometimes generate the number sequences without understanding the meaning of the numbers (Yilmaz, 2017). The same study suggested that teachers should assess students' number sense and design instructions with task-based interviews or activities to meet student needs. Considering the inadequate number sense development during early childhood education could be a reason for difficulty with mathematics even in adulthood (Jordan & Levine, 2009), the use of formative assessment in early childhood mathematics education becomes crucial.

Formative assessment is the pedagogical intersection of curriculum and assessment, aiming to boost learning instead of proving that learning occurs

(Crossouard & Pryor, 2012). Formative assessment not only has a positive influence on improving students' learning outcomes (Black & Wiliam, 1998; Kingston & Nash, 2011; Lee et al., 2020) but also it can boost students' performance on standardized assessments (Duckor et al., 2017; Reeves, 2001). While formative assessment and its implications have been examined (Atjonen, 2014; Elmahdi et al., 2018; Faber et al., 2016; Irving, 2015; Ramsey & Duffy, 2016), little is known about the relationship between formative and summative assessment in primary-grade mathematics classrooms.

Current Study

Considering formative and summative assessment are commonly used in educational settings and the conceptual difference in these two assessment paradigms, examining the relationship between formative and summative assessment became imperative due to multiple reasons. First, the relationship between formative and summative assessment could be an indication of effective formative assessment strategies. Second, this study could provide validity evidence for formative assessment tools. We used hierarchical linear modeling to examine if the relationship between formative and summative assessment change due to grade level and the average of students' achievement in a classroom.

The current study investigated the relationship between data from an internet-based formative assessment and data from a summative assessment focused on number sense for primary grade learners. The following research questions guided this study:

1. Is there a relationship between formative assessment scores and state summative assessment scores of students' mathematics number sense?
2. Is there a difference in the relationship between student's formative and summative assessment scores across kindergarten, first, and second grades?
3. Does the relationship between formative assessment scores and summative assessment scores vary by grade level and average students' achievement in a classroom?

Methods

Participants

Primary grade students' mathematics formative and summative assessment data came from a school district that included schools with urban,

suburban, and rural characteristics located in the southeastern United States. In this school district, 77 teachers participated in an 80-hours, one-year-long professional development experience, funded by the <state blinded> Mathematics Science Partnership grant program, to learn about formative assessment and students' number sense. Two groups of teachers with their students participated in this study. Group 1 consisted of 27 teachers and 258 students from Grade 1 (38%) and Grade 2 (62%). Group 2 consisted of 50 teachers and 477 students from Kindergarten (63%) and Grade 1 (37%). Group 1 and Group 2 took different parts of the online formative assessment tool (AMC Anywhere), which were hiding and counting, respectively.

Instruments

Internet-based Formative Assessment (AMC Anywhere)

The Internet-based formative assessment tool, AMC Anywhere, is used to collect data about students' mathematical understanding via diagnostic interviews in one-on-one settings (Figure 1). Based on the AMC Anywhere assessment tool, every student received a report with the letters A, P+, P, P-, I, and N (Figure 2). These letters, A, P, I, and N, stand for apply, practice, instruction, and needs prior skill, respectively. Please see Martin and Polly (2015) for a detailed description of the AMC Anywhere tool. In this project, teachers were expected to assess every student and use data to modify their classroom instruction. There are nine different assessments in the AMC Anywhere tool. Data from the Hiding and Counting Numbers Assessments were used in this study. The Hiding formative assessment provides data on students' ability to decompose numbers when they are given the total amount and one of the parts. For example, students are presented with a pile of 7 counters, they count them, and then the teacher hides some of them while the student looks away. The student sees the counters that remain and must determine how many counters are hiding and must orally explain their strategy. In the counting formative assessment, students count a set of objects put in front of them and determine how many counters there are and also determine how many would be in the pile if students added a counter or took away a counter. The Cronbach's alpha reliabilities for counting assessment scores and hiding assessment scores were .89 and .92, respectively, whereas the person reliability, which represents the probability of making the same separation between people over multiple measurements, was .92 for both of the assessments (Martin, Lambert et al., 2016). The consistent item location hierarchy provided evidence of the content aspect of the validity of the assessment scores with the developmental trajectory of children at these grade levels.

Figure 1
Hiding Assessment

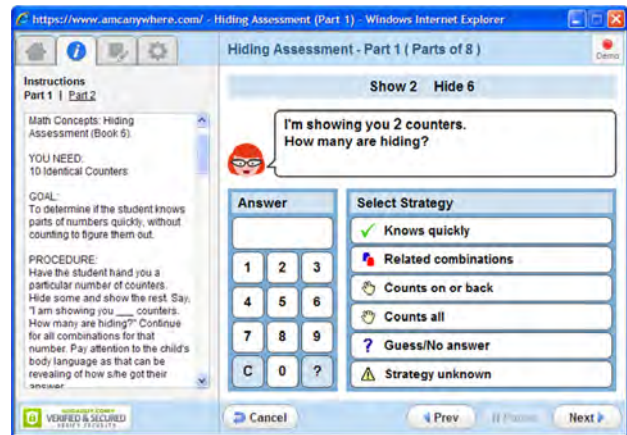


Figure 2
Hiding Assessment Scoring

Hiding Assessment - Part 1: Identifies Missing Parts of Numbers with Models								
Date	3	4	5	6	7	8	9	10
10/11/2013		A	P	I				
02/20/2014		A	P+	P				
05/22/2014			A	I				

Hiding Assessment - Part 2: Identifies Missing Parts of Numbers without Models							
Date	4	5	6	7	8	9	10
10/11/2013	N						
02/20/2014	P	P					
05/22/2014	P+	P	A	P	P	P	P

Summative Assessment

Teacher-leaders and state personnel created the summative assessment at the <state blinded> Department of Education. The assessment was designed to be given to an individual or small group of students in Kindergarten and Grade 1 and independently in Grade 2. Students' mathematics achievement was measured based on different tasks. The numbers of summative assessment tasks were 9, 12, and 11 in kindergarten, first grade, and second grade, respectively. Concepts that were assessed on those Kindergarten tasks included rote counting, counting objects, comparing numbers, addition, subtraction, and decomposing numbers. In first grade, summative assessment tasks included addition and subtraction, addition and subtraction word problems, solving for unknowns, extending the counting sequence, two-digit/place value, adding within 100, and subtracting multiples of 10. Lastly, second grade mathematics tasks included understanding place value, mentally adding 10 or 100, adding four two-digit numbers, addition and subtraction within 1000. The assessment included a rubric for each task. The rubrics described student performance on 3 possible levels: Level 1 (not yet meeting the standard), Level 2 (meeting the standards), and Level 3 (exceeding the standards). Students' summative assessment scores were the total percentage of tasks that students were grouped

at Level 2 or Level 3. Classroom achievement was measured as the percentage of proficient students in the classrooms.

Data Analysis

Item Response Theory was used to convert AMC Anywhere results into interval-level scale scores with a mean of 500 and a standard deviation of 100 with the Rasch Rating Scale Model (Andrich, 1978; Martin, Lambert et al., 2016). Pearson correlation was used to examine relationships between the two assessment scores by using .05 as the significance level. Before running the statistical analysis, the data were screened for outliers and normality. The multivariate outliers were checked based on Mahalanobis Distance, and the normality assumption was met.

For the first research questions, Pearson correlation coefficients were examined for the zero-order correlations between formative and summative test scores. The purpose of this analysis was that what percentage of the variance of the summative test scores was explained by the formative assessment scores. A one-way multivariate analysis of variance (MANOVA) was used to answer the second research question, which examined if there was a difference in students' formative and summative assessment scores between grade levels. Effect sizes were reported as eta squared (η^2), which is categorized as a small effect (0.01), medium effect (0.06), and large effect (0.15) (Cohen, 1988).

The third research question examined whether the relationship between formative and summative assessment data varied by the grade level and the average students' achievement in a classroom. We used a multi-level modeling procedure to examine the variance between classrooms relative to the variance within classrooms since students were nested in classrooms (Raudenbush & Bryk, 2002). Two-level hierarchical linear models (HLM) were used to understand the moderating effect of teachers on the relationship between the two student learning outcomes while grade level was controlled: formative assessment scores and state summative assessment

scores. Unconditional models were run first to calculate the intraclass correlation coefficients (ICC), and random intercept models were run later to check the moderating effect of grade level and teachers on the relationships. Group-mean centering was used for independent variables of formative assessments so that the intercept of the HLM Level I represents the expected classroom mean score on the summative assessment for a student whose formative assessment score was at the classroom means. The hiding and summative assessment unconditional model revealed an ICC value of .204, which indicated that 20.4% of the variance of summative assessment was between classrooms. Similarly, the counting and summative assessment unconditional model revealed an ICC value of .062, which indicated that 6.2% of the variance of summative assessment was between classrooms. The HLM were represented as:

Level 1:

$$Summative_{ij} = \beta_{0j} + \beta_{1j} * (Formative_{ij}) + r_{ij}$$

Level 2:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} * (Average Classroom Achievement_j) + \gamma_{02} * (Grade_{ij}) + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} * (Average Classroom Achievement_j) + \gamma_{12} * (Grade_{ij})$$

The same HLM models were used for counting and hiding assessment scores. Grade 1 and Grade 2 students took the Hiding formative assessment, so the grade was dummy coded, and Grade 1 was the base group. Similarly, students that took the counting formative assessment were in kindergarten and Grade 1, and kindergarten was the base group in this analysis.

Results

The Relationship between Formative and Summative Assessments

Descriptive statistics for the formative assessments and summative assessments were presented in Table 1.

Table 1
Descriptive Statistics of Hiding, Counting and Summative Assessment Scores

	Hiding			Counting				
	Grade	M	SD	n	Grade	M	SD	n
Hiding	1 st Grade	450.73	84.43	96	Kindergarten	288.18	40.18	302
	2 nd Grade	469.91	65.62	162	1 st Grade	315.59	11.73	169
	Total	462.77	73.61	258	Total	298.01	35.44	471
Summative	1 st Grade	87.69	19.11	96	Kindergarten	97.00	9.69	302
	2 nd Grade	77.04	24.06	162	1 st Grade	92.05	14.65	169
	Total	81.00	22.90	258	Total	95.23	11.93	471

Hiding assessment

The correlation between state summative assessment scores and hiding assessment scores was statistically significantly different from zero, $r = .44$ ($p < .01$). This meant that 19.36% of the variance of hiding assessment scores was explained by state summative assessment.

Counting assessment

After the removal of six outliers based on Mahalanobis distance, all assumptions for linear regression were met. The correlation between the state summative assessment scores and counting assessment scores was statistically significantly different from zero, $r = .29$ ($p < .01$). It indicated that 8.41% of the variance of counting assessment scores was explained by the state summative assessment scores.

The Difference in Formative and Summative Assessment Scores between Grade Levels

Hiding formative assessment

A MANOVA was conducted to examine the difference in students' hiding and state summative assessment scores based on grade level. The dependent variables were hiding, and summative assessment scores and the independent variable was grade levels as first grade and second grade.

The assumption of homogeneity covariance matrices was not satisfied (Box's $M = 18.32$, $F(3, 1413212) = 6.05$, $p < .001$). Thus, Pillai's Trace criterion was used, and the combined dependent variables were statistically different based on grade level, Pillai's Trace = .113, $F(2, 255) = 16.24$, $p < .001$, partial $\eta^2 = .11$ (medium effect size). Follow-up univariate F statistics showed statistically significant differences in hiding scores, $F(1, 256) = 4.14$, $p = .043$, partial $\eta^2 = .02$ (small effect size) and summative assessment scores, $F(1, 256) = 13.67$, $p < .001$, partial $\eta^2 = .05$ (small effect size).

Counting formative assessment

A MANOVA was conducted to examine the difference in students' counting and summative assessment scores based on grade level. The dependent variables were counting, and summative assessment scores, and the independent variable was grade levels as kindergarten and first grade.

The assumption of homogeneity covariance matrices was not satisfied (Box's $M = 311.55$, $F(3, 3735324) = 103$, $p < .001$), so Pillai's Trace criterion was used. The combined dependent variables were statistically different based on grade levels, Pillai's Trace = .23, $F(2, 468) = 70.31$, $p < .001$, partial $\eta^2 = .24$ (large effect size). Follow-up using univariate statistics suggested statistically significant differences in counting, $F(1, 469) = 75.05$, $p < .001$, partial $\eta^2 = .12$ (medium effect size) and summative assessment scores, $F(1, 469) = 19.42$, $p < .001$, partial $\eta^2 = .05$ (small effect size).

Moderating Effects of Classroom Achievement and Grade Levels

Parameter estimates of the intercepts and slopes in the HLMs were presented in Table 2.

Hiding formative assessment

The parameter estimates in Table 2 suggested that first grade students, whose hiding score was at the classroom mean, were expected to receive 80.45 out of 100 in their summative assessment. Grade-level differences in the hiding assessment scores was not statistically significant ($b = 3.73$, $p = .19$). This result indicated that second grade students' performance on the summative assessment was the same as the first graders' performance in terms of standardized scores. Classroom mean performance on the formative assessment, however, contributed to the expected classroom mean summative scores ($b = 0.81$, $p < .001$). With one unit increase in average classroom score on formative assessment, the summative assessment score at the classroom mean was expected to increase by 0.81 unit.

Table 2
Parameter Estimates of the Fixed Effects with Summative Assessment as the Outcome

	Level	Hiding			Counting		
		Estimate	SE	df	Estimate	SE	df
Intercept	2	80.41	1.66***	24	92.73	1.07***	47
Class Achievement	2	0.81	0.08***	24	0.43	0.10***	47
Grade	2	3.73	2.78	24	1.53	2.01	47
Formative Assessment	1	0.17	0.02***	228	0.21	0.04***	424
Grade	1	-0.002	0.03	228	0.38	0.18*	424
Class Achievement	1	-0.005	0.002**	228	-0.01	0.01	424

Note. * $p < .05$; ** $p < .01$; *** $p < .001$

The parameter estimates for the relationship between state summative assessment and hiding formative assessment at the individual level was statistically significant, $b = 0.17, p < .001$, which showed that with one unit increase in the student's hiding score from AMC Anywhere, the students' state summative assessment scores are expected to increase by 0.17. The parameter estimates for the moderating effect of classroom mean in formative assessment on this relationship was statistically significant, $b = -0.005, p = .01$, which demonstrated that with one unit increase in the average classroom assessment, the relationship between formative and summative assessments at the individual level was expected to decrease by 0.005 unit. The relationship was a bit weaker in classrooms with higher performance on hiding assessment. Grade level did not moderate this relationship, $b = -0.002, p = .94$, which meant that the relationship between formative assessment in hiding and summative assessment did not vary between first and second grade students. The effect size of this model was 30%, which indicated that independent variables included in the conditional model (hiding, class achievement, and grade) explained 30% of the individual differences in summative assessment. The conditional model also explained 94% of the between-classroom differences in summative assessment.

Counting formative assessment

The parameter estimates in Table 2 showed that kindergarten students whose counting scores were at the classroom mean were expected to receive 92.73 out of 100 in their summative assessment. Grade-level difference in counting assessment scores was not statistically significant ($b = 1.53, p = .45$). This result indicated that first grade students' performance on the summative assessment was the same as kindergarteners' performance in standardized scores. Classroom mean performance on formative assessment, however, contributed to the expected classroom mean summative scores ($b = 0.43, p < .001$). Data analysis indicated that a one-unit increase in the average classroom score on formative assessment was empirically associated with a 0.43 unit increase on the summative assessment score for the classroom mean.

The parameter estimates for the relationship between the summative assessment and the counting formative assessment at the individual level was statistically significant, $b = 0.21, p < .001$, which showed that with one unit increase in the student's counting score from AMC Anywhere, the students' state summative assessment scores were expected to increase by 0.21. Grade level moderated this relationship significantly, $b = 0.18, p < .05$, which means that the relationship between formative and summative assessment for counting is stronger

among first-graders than among kindergarteners. Classroom mean performance on counting did not moderate this relationship significantly, $b = -0.01, p = .19$. The effect size of this model is 27%, which means that independent variables that we included in the conditional model (counting, class achievement, and grade) explained 27% of the within-group variance of summative assessment at the individual level. The conditional model also explained 93% of the between-group variance of summative assessment.

Discussion

This study contributes to the literature in that the findings indicate an empirical relationship between the internet-based formative assessment, AMC Anywhere, and the summative assessment for primary grade learners. Similarly, Guo and Yan (2019) found that students' attitude toward formative assessment is a predictor of students' attitudes toward summative assessment. While it was known that formative assessment data could be a significant predictor of summative assessment for upper elementary and middle school students in mathematics (Golden, 2019; Steward, 2016), this study focused on primary grade learners and used multi-level modeling. Considering the existing literature that formative assessment can support student learning (Black & William, 1998; Kingston & Nash, 2011; Lee et al., 2020; Rakoczy et al., 2016), this finding indicates the benefit of formative assessment practices in mathematics classrooms as well as the external validity of the formative assessment instrument, AMC Anywhere formative assessment tool. Besides, the result could indicate alignment between formative and summative assessment (Golden, 2019). It is necessary to note that teachers should use the instructional tools correctly based on the principles for teaching to support student academic achievement (Chen et al., 2020)

Grade Level Difference in Formative Assessment and Summative Assessment

MANOVA results showed a significant difference between students' formative assessment and summative assessment results. Kindergarten students received higher formative assessment scores in hiding but lower summative assessment scores compared to first grade students. Similarly, the first grade students had higher formative assessment scores in counting but lower summative assessment scores compared to second grade students. In summary, lower grade students have higher formative assessment scores but lower summative assessment scores. This result echoed the Polly et al.'s (2018) finding that Kindergarten students outperformed First Grade students on formative assessment scores. Similarly, one explanation for increasing summative

assessment scores could be that students learn more mathematics content knowledge when they move up the grade levels (Martin, Polly et al., 2016; Martin, Lambert et al., 2016).

However, results from the multi-level model in this study indicated no statistically significant differences between Kindergarten and First Grade students in the formative assessments about counting and no statistically significant differences between first grade and second grade students in the formative assessment, hiding assessment, after moderating for classroom achievement. There could be multiple explanations of the different results from the MANOVA and the HLM analyses. First, the HLM did not compare the formative assessment and summative assessment between grade levels but instead used formative assessments to predict summative assessments. Second, HLM also used a different estimation approach (maximum likelihood estimation method) than that used in MANOVA (least squared estimation method). Third, as classroom achievement was included in the HLM model, it may explain the significant difference between formative and summative assessment between grade levels.

Moderating Effect of Classroom Achievement

The relationship between formative assessment in hiding and summative assessment is stronger in classrooms with lower average performance on formative assessment, which means that formative assessment may be more valuable for low achieving students. This result aligns with previous research as formative assessment was more effective for low-performing students (Polly et al., 2017; Bokhove & Drijvers, 2012; Koedinger et al., 2010). Students who were assessed with AMC Anywhere more frequently had a better understanding of number sense (Polly et al., 2017; 2018). Similarly, van den Berg et al. (2018) found that frequent use of formative assessment indicated higher achievement in fifth grades. The result of our study suggests that the frequency of formative assessment could be the reason for the difference between low performing and high performing students. It is important to interpret this result with caution because a possible explanation of this difference could be the faster growth of low achieving students (Polly et al., 2017).

Moderating Effect of Grade Level

The correlation between formative assessment in counting and summative assessment is higher in the first grade than in the kindergarten, which means that formative assessment could be more valuable for first grade students than for kindergarten students. This study aligned with findings from a previous study

(e.g., Polly et al., 2018) where Kindergarten students received higher scores on the formative assessment. Although grade level was not a significant moderator of formative assessment, task complexity could be a moderator for formative assessment (Kingston & Nash, 2011). Besides, Kluger and DeNisi (1996) found that the effect of feedback increases with task complexity. Since task complexity increases as grade level go up, first grade students may benefit more from formative assessment than kindergarten students due to task complexity.

This study analyzed the relationship between formative and summative assessment in primary-grade mathematics classrooms, but there are some limitations. First, student and teacher characteristics were not included in the analysis because demographic data were not available to the researchers. Second, as we concluded that formative assessment is more beneficial to low performing students, low performing student achievements may increase due to the ceiling effect. Therefore, we recommended that the researchers interpret results with caution.

Implications

This research provides insights into the relationship between formative assessment and summative assessment in primary grades mathematics classrooms. The literature suggested that formative assessment and data-based instructional changes could improve students' mathematics achievement (Fennema et al., 1996; Kingston & Nash, 2011; Lee et al., 2020). However, this study is unique in adding classroom achievement to multi-level analyses and examining grade level differences, and focusing on primary grades' learners. The results suggest that formative assessment can be more beneficial to encourage low achieving students in primary-grade mathematics classrooms. Therefore, we recommend teachers to use formative assessment practices more frequently in low achieved primary grade classrooms, which includes the cycle of data collection, data analysis, planning future instruction, and examining the impact of that instruction through cycling back to data collection. For instance, the cycle of AMC Anywhere tool is that formative assessment, instructional practices, and formative assessment, and AMC Anywhere uses technology to adjust formative assessment as students learn the concepts.

Teachers need to discover what aspect of number sense is challenging for low achieving students by using formative assessment. AMC Anywhere supports teachers in decision-making data-driven instructional decisions (Martin & Polly, 2015, p 376), and it collects data regarding students' strategies to solve the question. AMC Anywhere aims to reveal what students know or do not know about the number concepts rather

than whether students gave the right answer or not (Richardson, 2012). The positive relationship between formative and summative assessment may indicate the effectiveness of formative assessment strategies as the instruction quality improves student math achievement.

This study contributed to the literature as it examined the relationship between formative assessment and state summative assessment scores with both zero-order correlation coefficients and parameters estimated in a multi-level model by considering the nesting nature of students within classrooms. The significantly positive relationship between formative and summative assessment was noted in both methods. This relationship was stronger among first grade students than among kindergarteners. Further, low performing classrooms on formative assessment in hiding benefited more from formative assessment than high performing classrooms.

This study indicates the grade level difference between formative assessment and summative assessment in primary-grade mathematics classrooms. Thus, future studies should examine the grade level differences regarding formative and summative assessment for higher grades while including students' demographic information and teacher' background information. For instance, self-efficacy is a significant predictor of teachers' intention to use formative assessment (Karaman & Sahin, 2017). Further examination is also warranted to examine the moderating effect of grade level and students' proficiency level on the relationship between formative assessment and summative assessment. Results from these studies would be helpful for policymakers and educational practitioners who are interested in the use of formative assessment to guide instructions.

References

- American Educational Research Association, American Psychological Association and National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association, American Psychological Association and National Council on Measurement in Education.
- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43, 561-573.
- Atjonen, P. (2014). Teachers' views of their assessment practice. *Curriculum Journal*, 25(2), 238-259. Retrieved from <http://dx.doi.org/10.1080/09585176.2013.874952>
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74. doi:10.1080/0969595980050102
- Bokhove, C., & Drijvers, P. (2012). Effects of feedback in an online algebra intervention. *Technology, Knowledge and Learning*, 17(1-2), 43-59.
- Box, C., Skoog, G., & Dabbs, J. M. (2015). A case study of teacher personal practice assessment theories and complexities of implementing formative assessment. *American Educational Research Journal*, 52(5), 956-983. doi: 10.3102/0002831215587754
- Burke, K. (2010). *Balanced assessment: From formative to summative*. Bloomington, IN: Solution Tree Press.
- Carpenter, T., & Franke, M. (2004). Cognitively Guided Instruction: Challenging the Core of Educational Practice. In Glennan T., Bodilly S., Galegher J., & Kerr K. (Authors), *Expanding the Reach of Education Reforms: Perspectives from Leaders in the Scale-Up of Educational Interventions* (pp. 41-80). Santa Monica, CA; Arlington, VA; Pittsburgh, PA: RAND Corporation. Retrieved September 11, 2020, from <http://www.jstor.org/stable/10.7249/mg248ff.10>
- Chapman, O. (2012). Challenges in mathematics teacher education. *Journal of Mathematics Teacher Education*, 15, 263-270.
- Chen, I. H., Gamble, J. H., Lee, Z. H., & Fu, Q. L. (2020). Formative assessment with interactive whiteboards: A one-year longitudinal study of primary students' mathematical performance. *Computers & Education*, 150, 103833. <https://doi.org/10.1016/j.compedu.2020.103833>
- Cizek, G. J. (2010). An introduction to formative assessment: History, Characteristics, and Challenges. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 3-17). New York, NY: Routledge.
- Clark, I. (2010). Formative assessment: There is nothing so practical as a good theory. *Australian Journal of Education*, 54(3), 341-352. Retrieved from <http://dx.doi.org/10.1177/000494411005400308>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates Publishers.

- Connor, C. M., Mazzocco, M. M., Kurz, T., Crowe, E. C., Tighe, E. L., Wood, T. S., & Morrison, F. J. (2018). Using assessment to individualize early mathematics instruction. *Journal of School Psychology, 66*, 97-113.
- Crossouard, B., & Pryor, J. (2012). How theory matters: Formative assessment theory and practices and their different relations to education. *Studies in Philosophy and Education, 31*(3), 251-263. doi: 10.1007/s11217-012-9296-5
- Dixson, D. D. & Worrell, F. C. (2016). Formative and summative assessment in the classroom. *Theory Into Practice, 55*(2), 153-159, DOI: 10.1080/00405841.2016.1148989
- Duckor, B., Holmberg, C., & Becker, J. R. (2017). Making moves: Formative assessment in mathematics. *Mathematics Teaching in the Middle School, 22*, 334-342. Retrieved from <http://about.jstor.org>
- Elmahdi, I., Al-Hattami, A., & Fawzi, H. (2018). Using technology for formative assessment to improve students' learning. *Turkish Online Journal of Educational Technology-TOJET, 17*(2), 182-188.
- Faber, J. M., Luyten, H., & Visscher, A. J. (2016). The effects of a digital formative assessment tool on mathematics achievement and student motivation: Results of a randomized experiment. *Computers & Education, 106*, 83-96.
- Fennema, E., Franke, M. L., Carpenter, T. P., & Carey, D. A. (1993). Using children's mathematical knowledge in instruction. *American educational research journal, 30*(3), 555-583.
- Fennema, E., Carpenter, T. P., Franke, M. L., Levi, L., Jacobs, V. R., & Empson, S. B. (1996). A longitudinal study of learning to use children's thinking in mathematics instruction. *Journal for research in mathematics education, 40*3-434.
- Gersten, R., & Chard, D. (1999). Number sense: Rethinking arithmetic instruction for students with mathematical disabilities. *Journal of Special Education, 33*(1), 18-28.
- Golden, J. B. (2019). *The Relationship between Formative Math Assessments and the State Summative Assessment for Middle School Students in Alabama* (Doctoral dissertation, University of South Alabama). (Order No. 27544015). Available from ProQuest Dissertations & Theses Global. (2316839127).
- Govender, P. (2019). Formative assessment as 'formative pedagogy' in grade 3 mathematics. *South African Journal of Childhood Education, 9*(1), 1-12.
- Guo, W. Y., & Yan, Z. (2019). Formative and summative assessment in Hong Kong primary schools: students' attitudes matter. *Assessment in Education: Principles, Policy & Practice, 26*(6), 675-699. DOI: 10.1080/0969594X.2019.1571993
- Hattie, J. (2003). Formative and summative interpretations of assessment information. Retrieved September 28, 2011.
- Heritage, M. (2007). Formative assessment: What do teachers need to know and do? *Phi Delta Kappan, 89*(2), 140-145.
- Heritage, M. (2013). *Formative assessment in practice: A process of inquiry and action*. Boston: Harvard Education Press.
- Irving, K. (2015). Technology-assisted formative assessment. In M. J. Urban & D. A. Falvo (Eds.), *Improving K-12 STEM education outcomes through technological integration* (380 - 398). <http://doi.org/10.4018/978-1-4666-9616-7.ch017>
- Jordan, N. C., & Levine, S. C. (2009). Socioeconomic variation, number competence, and mathematics learning difficulties in young children. *Developmental Disabilities Research Reviews, 15*(1), 60-68.
- Kanjee, A., & Sayed, Y. (2013). Assessment policy in post-apartheid South Africa: Challenges for improving education quality and learning. *Assessment in Education: Principles, Policy & Practice, 20*(4), 442-469.
- Karaman, P., & Şahin, Ç. (2017). Adaptation of Teachers' Conceptions and Practices of Formative Assessment Scale into Turkish Culture and a Structural Equation Modeling. *International Electronic Journal of Elementary Education, 10*(2), 185-194. DOI: 10.26822/iejee.2017236114
- Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice, 30*(4), 28-37.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254- 284.

- Koedinger, K. R., McLaughlin, E. A., & Heffernan, N. T. (2010). A quasi-experimental evaluation of an online formative assessment and tutoring system. *Journal of Educational Computing Research, 43*(4), 489-510.
- Lee, H., Chung, H. Q., Zhang, Y., Abedi, J., & Warschauer, M. (2020). The Effectiveness and Features of Formative Assessment in US K-12 Education: A Systematic Review. *Applied Measurement in Education, 33*(2), 124-140. DOI: 10.1080/08957347.2020.1732383
- Martin, C., Lambert, R., Polly, D., Wang, C., & Pugalee, D. (2016). The measurement properties of the assessing math concept's assessments of primary students' number sense skills. *Journal of Applied Measurement, 17*, 354-364.
- Martin, C. S., & Polly, D. (2015). Using the AMC Anywhere web-based assessment system to examine primary students' understanding of number sense. In *Cases on technology integration in mathematics education* (pp. 365-377). IGI Global.
- Martin, C. S., Polly, D., Wang, C., Lambert, R. G., & Pugalee, D. K. (2016). Perspectives and practices of elementary teachers using an Internet-based formative assessment tool: The case of assessing mathematics concepts. *International Journal for Technology in Mathematics Education, 23*(1), 3-13.
- Mason, B. J., & Bruning, R. (2001). Providing feedback in computer-based instruction: What the research tells us. Retrieved February 19, 2019.
- McMillan, J. H. (2010). The practical implications of educational aims and contexts for formative assessment. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 41-58). New York, NY: Routledge.
- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA.
- National Council of Teachers of Mathematics (NCTM). (2014). Principles to actions: Ensuring mathematical success for all.
- National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards for Mathematics*. Washington, DC: Authors.
- North Carolina Department of Public Instruction. (2017). Standards for mathematical practice. Retrieved from <https://files.nc.gov/dpi/documents/curriculum/mathematics/scos/current/k-2.pdf>
- Pachler, N., Daly, C., Mor, Y., & Mellar, H. (2010). Formative e-assessment: Practitioner cases. *Computers & Education, 54*(3), 715-721.
- Perie, M., Marion, S., & Gong, B. (2009). Moving toward a comprehensive assessment system: A framework for considering interim assessments. *Educational Measurement: Issues and Practice, 28*(3), 5-13.
- Polly, D., Martin, C. S., Wang, C., Lambert, R. G., & Pugalee, D. K. (2016). Primary grades teachers' instructional decisions during online mathematics professional development activities. *Early Childhood Education Journal, 44*, 275-287. doi: 10.1007/s10643-015-0711-8
- Polly, D., Wang, C., Martin, C., Lambert, R., Pugalee, D., & Middleton, C. (2018). The influence of Mathematics professional development, school-level, and teacher-level variables on primary students' Mathematics achievement. *Early Childhood Education Journal, 46*, 31-45. doi: 10.1007/s10643-017-0837-y
- Polly, D., Wang, C., Martin, C., Lambert, R. G., Pugalee, D. K., & Middleton, C. W. (2017). The influence of an internet-based formative assessment tool on primary grades students' number sense achievement. *School Science and Mathematics, 117*(3-4), 127-136.
- Rakoczy, K., Pinger, P., Hochweber, J., Klieme, E., Schütze, B., & Besser, M. (2019). Formative assessment in mathematics: Mediated by feedback's perceived usefulness and students' self-efficacy. *Learning and Instruction, 60*, 154-165.
- Ramsey, B., & Duffy, A. (2016). *Formative assessment in the classroom: Findings from three districts. Michael and Susan Dell Foundation and Education, 1*. Retrieved from <https://education-first.com/wp-content/uploads/2016/05/MSDF-Formative-Assessment-Study-Final-Report.pdf>
- Randel, B., Aphorhp, H., Beesley, A. D., Clark, T. F., & Wang, X. (2016). Impacts of professional development in classroom assessment on teacher and student outcomes. *The Journal of Educational Research, 109*(5), 491-502. doi:10.1080/00220671.2014.992581

- Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.
- Reeves, D. B. (2001). Standards make a difference: The influence of standards on classroom assessment. *NASSP Bulletin*, 85(621), 5-12.
- Richardson, K. (2012). *How children learn number concepts: A guide to the critical learning phases*. Bellingham, WA: Math Perspectives.
- Riggan, M., & Olah, L. N. (2011). Locating interim assessments within teachers' assessment practice. *Educational Assessment*, 16(1), 1-14.
- Shepard, L. A. (2008). Formative assessment: Caveat emptor. In C. A. Dwyer (Ed.). *The future of assessment: Shaping teaching and learning*. New York, NY: Erlbaum.
- Shepard, L. A., Penuel, W. R., & Pellegrino, J. W. (2018). Using learning and motivation theories to coherently link formative assessment, grading practices, and large-scale assessment. *Educational Measurement: Issues and Practice*, 37(1), 21-34.
- Stewart, A. B. B. (2016). *An analysis of the predictive validity of curriculum-based mathematics assessments on student performance on state standardized tests*. Available from ProQuest Dissertations & Theses Global. (AAI3704816) Retrieved from <https://search.proquest.com>
- Thames, M. H., & Ball, D. L. (2010). What math knowledge does teaching require? *Teaching Children Mathematics*, 17, 220-229.
- van den Berg, M., Bosker, R. J., & Suhre, C. J. (2018). Testing the effectiveness of classroom formative assessment in Dutch primary mathematics education. *School Effectiveness and School Improvement*, 29(3), 339-361. <https://doi.org/10.1080/09243453.2017.1406376>
- Veldhuis, M., & van den Heuvel-Panhuizen, M. (2019). Supporting primary school teachers' classroom assessment in mathematics education: Effects on student achievement. *Mathematics Education Research Journal*, 1-23.
- Yilmaz, Z. (2017). Young children's number sense development: Age related complexity across cases of three children. *International Electronic Journal of Elementary Education*, 9(4), 891-902.