Original Research Article

Journal of Special Education Technology
2025, Vol. 0(0) 1–13

DOI: 10.1177/01626434251314041
journals.sagepub.com/home/jst
Sage

Digital Pencil Usage and Mathematics Performance Among Students with Learning Disabilities and Their General Education Peers

Xin Wei[1], Susu Zhang[2], and Jihong Zhang[3]

Abstract

This investigation explores the relationship between the use of digital pencil and mathematical problem-solving accuracy among 1,530 students with learning disabilities (LD) and 25,400 general education (GE) peers from the 2017 digital National Assessment of Educational Progress mathematics assessment. The term "digital pencil" in this context refers to NAEP's "embedded pencil," a scratchwork tool within the digital assessment interface that allows students to draw or annotate on the screen using either a stylus or their finger. Findings reveal that students with LD utilized digital pencils less frequently than their GE peers, particularly on more complex items. However, digital pencil use was associated with a 20% increase in the likelihood of GE students accurately solving difficult problems and a 26% increase in accuracy for students with LD solving simpler problems. The study highlights the educational implications of incorporating digital tools like the digital pencil in learning and assessment environments, emphasizing the need for tailored instructional strategies to support diverse learners.

Keywords

digital pencil usage, mathematics performance, task difficulty, learning disabilities, NAEP

Introduction

The acquisition of core mathematical competencies is pivotal for academic achievement and career readiness. Data from the National Assessment of Educational Progress (NAEP) reveal a troubling trend: the achievement gaps between students with disabilities and their general education (GE) peers are widening, with a marked decline in proficiency rates for students with disabilities (NAEP, 2022). The exacerbation of these disparities due to the pandemic underscores an urgent need to fortify the mathematical capabilities of students with disabilities.

Students with learning disabilities (LD) encompass a significant demographic within special education and face distinctive challenges in mathematical learning. These challenges span computational difficulties, memory constraints, conceptual misunderstandings, difficulties in selecting appropriate strategies, and self-regulation issues during problem-solving (Geary, 2004; Grigorenko et al., 2020). These impediments often manifest in various stages of problem-solving, including comprehension, interpretation, planning, and execution (Krawec et al., 2012; Mayer, 1985). Addressing these challenges requires innovative educational tools and strategies that cater to their unique learning needs, and this study explores the use of the digital pencil, specifically referring to NAEP's "embedded pencil" feature, which allows students to draw and annotate directly on the screen during the assessment.

Drawing as a Learning Strategy

Drawing as a learning strategy enables students to create pictorial representations to effectively grasp and retain instructional content. Quillin and Thomas (2015) describe it as a method whereby learners produce drawings to better understand textual materials, thereby enhancing learning outcomes. This strategy has been supported by nearly five decades of research highlighting its benefits as an educational tool (van Meter & Garner, 2005). For example, Schwamborn et al. (2010) showed that participants who engaged in drawing while learning about the chemistry of laundry retained and transferred knowledge more effectively

[1]Digital Promise, Washington, DC, USA
[2]University of Illinois, Urbana-Champaign, Champaign, IL, USA
[3]University of Arkansas, Fayetteville, AR, USA

Corresponding Author:
Xin Wei, Learning Science Research, Digital Promise, 1001 Connecticut Avenue NW, Suite 935, Washington, DC 20036, USA.
Email: xwei@digitalpromise.org

than those who solely relied on reading the material. Similarly, Wammes et al. (2018) found that drawing aided in the better retention of definitions compared to merely writing them down.

The efficacy of drawing as a learning tool may stem from its engagement of both cognitive functions, such as selecting and organizing information, and metacognitive functions, including monitoring accuracy. These processes collectively foster a deeper engagement with learning materials (van Meter & Firetto, 2013). In the realm of mathematics education, drawing visual representations is strongly recommended to help students, particularly those with LD, understand abstract concepts and solve problems more effectively (Griffin et al., 1994; NCTM, 2000; NRC, 2001). Studies have demonstrated that students with LD who are taught to use diagrams experience significant gains in mathematics performance compared to their peers receiving traditional instruction (van Garderen et al., 2012; Garderen et al., 2014). Additionally, there is growing interest in how students spontaneously generate drawings on scratch paper during problem-solving. Research in this area suggests that students who create accurate schematic representations typically achieve better outcomes in mathematics (Boonen et al., 2014; Krawec, 2014).

Digital Pencils as Educational Aids

With the rise of touchscreen technology, touchscreen tablets and digital pencils are popular digital mediums for drawing activities that can achieve similar operability as traditional pencil drawing (Wammes et al., 2018). Digital pencils have become an indispensable tool in digital learning and assessments. These tools enhance the learning experience by offering dynamic and interactive platforms that support visual representation, annotation, and problem-solving, helping to clarify abstract mathematical concepts and enable students to visualize their thoughts through diagrams, graphs, and sketches (Rau, 2017; Xie & Zhou, 2024). Studies have found evidence of digital pencils aid in memory retention and boost engagement (Wammes et al., 2018). For instance, Lee and Cheng (2021) demonstrated how digital pencils could enhance fourth graders' understanding and motivation in learning about colors. The utility of digital drawing tools extends significantly to special education and students who typically achieve lower academic results. Rubin et al. (2015) observed that the use of digital pencils helped fourth and fifth graders in special education classes improve their multiplication skills. Similarly, Patti and Garland (2015) reported improvements in study skills, independent work, and assessment accommodations for students with LD who used digital pencils. Further emphasizing their versatility, Kwak and Gweon (2019) found that digital scratchpads could significantly enhance arithmetic problem-solving capabilities, particularly among students who traditionally perform at lower academic levels.

Addressing the Research Gap

Despite promising evidence, research exploring the effectiveness of digital pencils has been limited in scope, often relying on small, non-representative samples. There remains a lack of studies using nationally samples to examine the usage pattern of digital pencil and its association with performance. Furthermore, the existing literature seldom examines the differential impact of digital pencils on students with LD versus their GE peers and does not thoroughly explore how task difficulty and disability status might interact to influence the efficacy of digital pencils (Kwak & Gweon, 2019). This study aims to fill this gap by leveraging data from NAEP to explore the digital pencil usage pattern and its association with mathematical problem-solving accuracy across task difficulty and disability status. The following research questions guide this inquiry:

1. Do students with LD differ from their GE peers in the usage of digital pencils for mathematical problem-solving?
2. What is the relationship between the use of digital pencils and the accuracy of mathematical problem-solving, and does it vary with the difficulty of the task?
3. How does the intersection of task difficulty and LD status affect the efficacy of digital pencils in mathematical problem-solving?

By addressing these questions, the study seeks to inform the design of digital learning and assessment systems that accommodate the diverse needs of all learners, particularly those with learning disabilities.

Methods

Data

The NAEP represents the largest nationally representative and continuing assessment of what America's students know and can do in various subject areas. For the 2017 NAEP mathematics assessment, 144,900 eighth graders from 6,500 schools across the United States were sampled using a deeply stratified multistage cluster sampling technique. The NAEP mathematics assessment was conducted using Microsoft Surface Pro tablets, which were equipped with external keyboards, styluses, and earbuds for audio accommodations. At the beginning of the assessment, students participated in an interactive tutorial to familiarize themselves with the system's features, including the digital pencil (also referred to as the embedded pencil). Each student participant was randomly assigned two 30-min blocks math items from a pool of 10 different blocks. After completing the math assessment, students completed a subsequent 15-min demographic survey, all administered via digital tablets. These blocks were administered consecutively on the same day.

Study Sample

The NAEP publicly released process data for one block of math items from the 2017 NAEP mathematics assessment, out of a total of 10 blocks administered. While the full NAEP dataset is nationally representative, this single block of process data is not. Despite its lack of national representativeness, this block offers valuable insights into student behavior and performance during the assessment. The process data provide detailed logs of student interactions with the assessment interface, including the use of the digital pencil, along with student responses to each test item, demographic information, and survey data. For our analysis, we used the entire publicly released block, which includes data from 1,530 students with LD[1], 1,270 students with other disabilities, and 25,400 GE peers from about 270 schools. However, students with disabilities other than LD were excluded from this analysis, as they were not the focus of this study.

For each student with disabilities who was sampled to participate in the NAEP test, a principal/assistant principal, special education teacher, bilingual education/ESL teacher, or classroom teacher filled out a disability questionnaire for that student (NAEP, 2020). This questionnaire collected information about the student's disability category, severity of the disability, grade level, whether a student with disabilities needs test accommodations, and what accommodation a student should receive on the NAEP test. Students with LD in the sample were identified by their schools as receiving special education services under the specific learning disability category.

Measures

Student Demographic Characteristics. Demographic variables include gender, age at testing, race/ethnicity (coded into four groups: African American, Hispanic, White, and Other), and eligibility for free or reduced-price lunch. Descriptive statistics of these characteristics are tabulated in Table 1 for students with LD and GE group.

To assess the representativeness of our sample, we compared the demographic characteristics of students in our study to national averages reported by the National Center for Education Statistics (NCES, 2024). The percentage of students with LD in our sample (approximately 5.2%) closely aligns with national data, where 15% of students are identified as having disabilities, and 32% of these students are classified as having learning disabilities (NCES, 2024).

NAEP Universal Design Digital Pencil. The NAEP employs universal design principles in its digital assessments to accommodate a wide range of abilities and learning styles (Rose et al., 2016). This digital pencil is part of the scratchwork tools available in the test interface to all students. By selecting the scratchwork icon, students can draw, annotate, or write directly on the test screen. They are given the option to use either a stylus or their finger to interact with the screen, offering flexibility in how they make notations.

This tool is posited to alleviate cognitive load, particularly on complex items, by facilitating direct on-screen manipulation and response (Way & Strain-Seymour, 2021). For each item, this study coded "digital pencil" as 1 if a student used the digital pencil or 0 if a student did not use the digital pencil. Please note that the "digital pencil" variable is an item-varying variable because students choose whether to use the digital pencil for each item. However, the process data do not capture how the digital pencil was used — whether for drawing mathematical representations, performing calculations, or making unrelated marks. This limitation constrains our ability to assess the exact role the digital pencil played in problem-solving.

NAEP digital assessments did not automatically provide all students with scratch paper and pencil. But if students specifically requested these materials during the administration, they were permitted to use scratch paper and pencils.

Table 1. Demographic Characteristics of Students with Learning Disabilities and Their General Education Peers.

Variables	Learning disabilities	General education	Overall
Male, %	62***	49	50
White, %	49	52	52
African American, %	18	16	16
Hispanic, %	24	21	21
Other, %	9	11	11
Free or reduced lunch status, %	63***	47	48
Age, Mean (SD)	14.6 (0.6)	14.4 (0.5)	14.4 (0.5)

***$p < .001$.
SD = standard deviations. Chi-square test results in the LD column indicates the significant differences between students with learning disabilities and their general education peers.
Note. Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), Response Process Data From the 2017 NAEP Grade 8 Mathematics Assessment. Samples sizes were rounded to the nearest 10 following NAEP restricted data use agreement.

Table 2. Item Content and Easiness, Digital Pencil Usage, and Mathematics Problem Solving Accuracy by Group and Mathematics Item.

Item	Content	Easiness[a]	Sample	% used digital pencil	% correct	% correct w/ Digital pencil	% correct w/o digital pencil
1	Translate a percent to a fraction	1.7	Whole sample	5.9%	64.3%	59.8%***	64.6%
			LD	6.1%	30.2%	24.5%	30.6%
			GE	5.9%	66.3%	62.1%***	66.6%
2	Complete a circle graph to represent data	4.2	Whole sample	6.7%	94.6%	93.7%	94.6%
			LD	6.7%	88.4%	92.1%	88.2%
			GE	6.7%	94.9%	93.8%	95.0%
3	Multiplication of two two-digit decimals	0.8	Whole sample	28.8%	46.4%	46.5%	46.4%
			LD	28.2%	19.1%	23.5%**	17.4%
			GE	28.8%	48.0%	47.8%	48.1%
4	Determine x and y intercept of a given line	0.9	Whole sample	7.9%	47.7%	43.3%***	48.1%
			LD	8.0%	22.5%	27.1%	22.2%
			GE	7.9%	49.2%	44.2%***	49.6%
5	Compare measurement using unit conversions	1.4	Whole sample	10.1%	59.2%	62.1%***	58.9%
			LD	6.8%***	34.5%	44.2%*	33.7%
			GE	10.3%	60.7%	62.9%**	60.5%
6	Extend a numerical pattern	0.7	Whole sample	2.6%	44.2%	46.2%	44.1%
			LD	2.3%	15.1%	32.4%***	14.7%
			GE	2.6%	45.9%	46.9%	45.9%
7	Calculate diameter of a circle from a given circumference	−1.3	Whole sample	20.6%	12.2%	8.5%***	13.2%
			LD	16.1%***	14.0%	5.3%***	15.7%
			GE	20.8%	12.1%	8.6%***	13.1%
8	Rotation of a triangle	0.2	Whole sample	6.6%	36.9%	36.6%	36.9%
			LD	6.2%	24.0%	22.3%	24.1%
			GE	6.7%	37.7%	37.4%	37.7%
9	Create a proportion to find distance on a map	1.8	Whole sample	6.2%	67.4%	65.4%*	67.5%
			LD	6.1%	42.1%	55.0%**	41.3%
			GE	6.2%	68.9%	66.0%*	69.1%
10	Identify characteristics of lines	−1.4	Whole sample	7.5%	12.8%	23.9%***	11.9%
			LD	3.5%***	2.4%	5.7%	2.3%
			GE	7.7%	13.5%	24.4%***	12.6%
11	Make and explain a conclusion about linear equations	−1.2	Whole sample	5.6%	14.7%	21.3%***	14.3%
			LD	3.9%**	2.3%	5.0%	2.2%
			GE	5.7%	15.5%	21.9%***	15.1%
12	Identify figures that are composites of 2 given shapes	−2.2	Whole sample	9.9%	6.8%	10.5%***	6.3%
			LD	5.9%***	1.2%	0%	1.3%
			GE	10.2%	7.1%	10.9%***	6.7%
13	Evaluate circle graph and bar graph to determine possible data sets	−2.4	Whole sample	5.4%	5.9%	10.8%***	5.6%
			LD	3.1%***	0.6%	0%	0.6%
			GE	5.6%	6.2%	11.1%***	5.9%
14	Match box-plots to stem-and-leaf plots	−0.7	Whole sample	4.2%	20.0%	28.0%***	19.6%
			LD	3.0%*	9.4%	6.7%	9.5%
			GE	4.3%	20.6%	28.9%***	20.3%
15	Write expression for polygon area using conjecture	−1.8	Whole sample	11.4%	9.4%	14.7%***	8.6%
			LD	9.0%**	0.5%	0.8%	0.5%
			GE	11.5%	9.9%	15.4%***	9.1%

Note. % = percent; LD = students with learning disabilities; GE = general education students; % Correct w/ Pencil = the percentage of correct responses from students who utilized the digital pencil for this test item; % Correct w/o Pencil = the percentage of correct answers from students who did not use the digital pencil for this test item.
Significance notations in column "% Used Digital Pencil" are two-proportion z-test results that were used to test whether there was a significant difference in the proportion of students using the digital pencil on a item between students with LD and GE peers.
Significance notations in column "% Correct w/Pencil" are two-proportion z-test results that were used to test whether there was a significant difference in the proportion of students answered an item correctly between students used the digital pencil and students who did not use the digital pencil.
Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress, Response Process Data From the 2017 NAEP Grade 8 Mathematics Assessment. All 15 items can be found at https://nces.ed.gov/NationsReportCard/nqt/Search
$^*p < .05$, $^{**}p < .01$, $^{***}p < .001$.
[a]"Item easiness" is modelled as a random intercept component in mixed model.

Item Performance. The mathematics assessment includes a total of 15 test items, each described in Table 2. These items encompass topics such as fractions, lines, shapes and rotations, the product of two two-digit decimals, x- and y-intercepts, circle graphs, box plots, stem-and-leaf plots, the diameter and circumference of circles, and the area of shapes. The items in each NAEP mathematics block were administered in a fixed, non-random order to all students. As shown in Table 2, items are arranged by difficulty, with more challenging items typically appearing toward the end of the session. This fixed ordering may have influenced students' use of the digital pencil, particularly if they began to run out of time or feel fatigued. In such cases, students may have rushed through or skipped the final questions, which could explain the lower frequency of digital pencil use and a potential decline in accuracy on the more difficult items.

Six items have a maximum score of 1: the student scored 1 if she/he answered the item correctly or 0 if she/he answered the item incorrectly. Eight items have a maximum score of 2, and one item has a maximum score of 4: Incorrect responses are assigned the score of 0 and correct responses are assigned the maximum score, with partially correct responses scored in between. We recoded scores on these nine items into binary scores, with 1 for correct and 0 for partially correct or incorrect.

Statistical Analysis

Descriptive Analysis. All analyses were conducted using R version 4.1.0 (R Core Team, 2021). Initial descriptive analyses explored demographic differences between students with LD and their GE peers. We then analyzed digital pencil usage and performance, broken down by item and group membership. Two-proportion z-tests were employed to assess: (1) the differences in digital pencil usage between the LD and GE groups for each item, and (2) the within-group differences in item correctness between students who used the digital pencil and those who did not.

Generalized Linear Mixed Models (GLMMs)

Generalized Linear Mixed Models (GLMMs) have been widely used in psychophysical (e.g., Moscatelli et al., 2012), medical (e.g., Stijnen et al., 2010), and statistical (e.g., Booth & Hobert, 1999) fields. Compared to other psychometric models, GLMMs allow researchers to incorporate the multilevel structure into the measurement model. The psychometric community has used GLMMs to estimate test item difficulty and examinee ability using the item response theory (e.g, Wang et al., 2022; Boeck et al., 2011). A GLMM estimates fixed and random effects when dependent variables are not normally distributed (Boeck et al., 2011).

This study used GLMMs to fit a 1-Parameter Logistic (1PL) item response theory (IRT) model with random person and item effects (Boeck et al., 2011). In this model, the probability of a correct response is a function of both the individual's latent trait and item difficulty. The 1PL model assumes all items to be equally discriminating and guessing to be absent; therefore, the only item parameter is the difficulty level of the item. Specifically, for item i and person p, the logit of the correct response probability, η_{pi}, is

$$\eta_{pi} = \beta_0 + b_{0p} + b_{0i}, \tag{1}$$

where β_0 is the overall intercept across items and persons, b_{0p} is the person random intercept indicating person ability, and b_{0i} is the item random intercept indicating item easiness. This is the base model – Model 1.

To answer RQ 2, we added an item-level binary digital pencil variable, DP_{pi}, to the above 1PL IRT model to get Model 2:

$$\eta_{pi} = \beta_0 + b_{0p} + b_{0i} + (\beta_1 + b_{1i})DP_{pi} \tag{2}$$

DP_{pi} takes the value of 1 if student p used the digital pencil on item i and 0 otherwise. The overall draw effect is hence represented by the fixed effect β_1. The random slope for the draw variable is the random draw effects variable across items—b_{1i} indicates how the fixed effect β_1 is adjusted by item. The by-item adjustment b_{1i} and item intercept b_{0i} are tied to the same item; thus, their correlation can be tested to examine whether the impact of draw depends linearly on item difficulty. b_{0p} denotes the random intercepts which represents students' latent ability.

To answer RQ 3, we added a binary learning disability indicator variable, LD_p, and its interaction with digital pencil, DP_{pi}, to the above Model 2 to get Model 3:

$$\eta_{pi} = \beta_0 + b_{0p} + b_{0i} + (\beta_1 + b_{1i})DP_{pi} + (\beta_2 + b_{2i})LD_p + (\beta_3 + b_{3i})DP_{pi}LD_p \tag{3}$$

In addition to the parameters described for Model 2, Model 3 includes a new student-level variable, LD_p, which equals 1 if student p belongs to the LD group and 0 otherwise. The overall LD effect on correct response probability is represented by the fixed effect β_2. The random LD effects variable across items, b_{2i}, indicates how the fixed effect β_2 is adjusted by LD. A DP-by-LD interaction term was added to Model 3. The overall DP-by-LD interaction effect is represented by the fixed effect β_3. The random draw by LD effects, b_{3i}, indicates how the draw-by-LD effect varies by item. For each item i, there are four item random effects: the random intercept (b_{0i}), the random slope for the main effect of using digital pencil (b_{1i}), the random slope for the main effect of LD status (b_{2i}), and the random slope for the DP-by-LD interaction (b_{3i}). The GLMM analysis presents a four-by-four correlation matrix for the four random effects. The correlation between b_{0i} and b_{3i}, $Corr(b_{0i}, b_{3i})$, was used to test whether the interaction of draw by LD depends linearly on item difficulty.

Given the significant differences in gender and FRPL status between the LD and GE groups, we included these variables as covariates in our generalized linear mixed models to account for their potential confounding effects on the relationship between digital pencil usage and mathematics performance. Controlling for these demographic factors allows for a more precise estimate of the effect of digital pencil usage in our Model 4.

The GLMMs were fitted using the R lme4 package (Boeck et al., 2011). Model fit was evaluated using information criteria (AIC and BIC) and log-likelihood ratios to determine if the inclusion of additional parameters significantly improved model fit (Vrieze, 2012).

Visualization of Effects. We plotted the effect sizes of the digital pencil against item difficulty. This approach helps to clarify whether the tool's effectiveness is consistent across different levels of item complexity and between different student groups. Specifically, Figure 1 presents the effect size of using digital pencil on each item, ranking item from the hardest to the easiest, for the whole sample. Each dot in Figure 1 represents an item. Effect size is calculated as the sum of overall digital pencil fixed effect across all items and the random digital pencil effect for a particular item, that is,

$$\beta_{1i} = \beta_1 + b_{1i}.$$

It indicates the increase in log-odds of responding correctly on item i if the student used the pencil on that item. A positive effect size indicates the group of students who used the digital pencil had higher odds of answering that item correctly. For example, an effect size of 0.2, is equivalent to an increased odds in answering an item correctly by 22% ($e^{0.2} - 1 = 22\%$) if the student had used the digital pencil.

We also plotted the relationship between item easiness and the effect sizes of digital pencil on each item for LD group versus GE group in Figure 2. Specifically, using the fixed and random effect estimates from Model 3 (see Equation (3)), the effect size of using digital pencil on correct response probability on item i for the GE group β_{1i} is given by

$$\beta_{1i} = \beta_1 + b_{1i},$$

and the effect size of digital pencil for the LD group β_{3i} is given by

$$\beta_{3i} = \beta_1 + b_{1i} + \beta_3 + b_{3i}.$$

Results

Descriptive Analysis Results

Our analysis in Table 1 revealed significant demographic differences between the two groups. Specifically, students with LD group included a higher percentage of male students (62% compared to 49% in the GE group; $p < .001$) and a higher proportion of students receiving free or reduced-price lunch (FRPL) (63% compared to 47% in the GE group; $p < .001$) than the GE group. To account for potential confounding effects, we included these two variables in our GLMM analysis (Model 4).

Our descriptive findings, presented in Table 2, reveal notable disparities in the utilization of the digital pencil

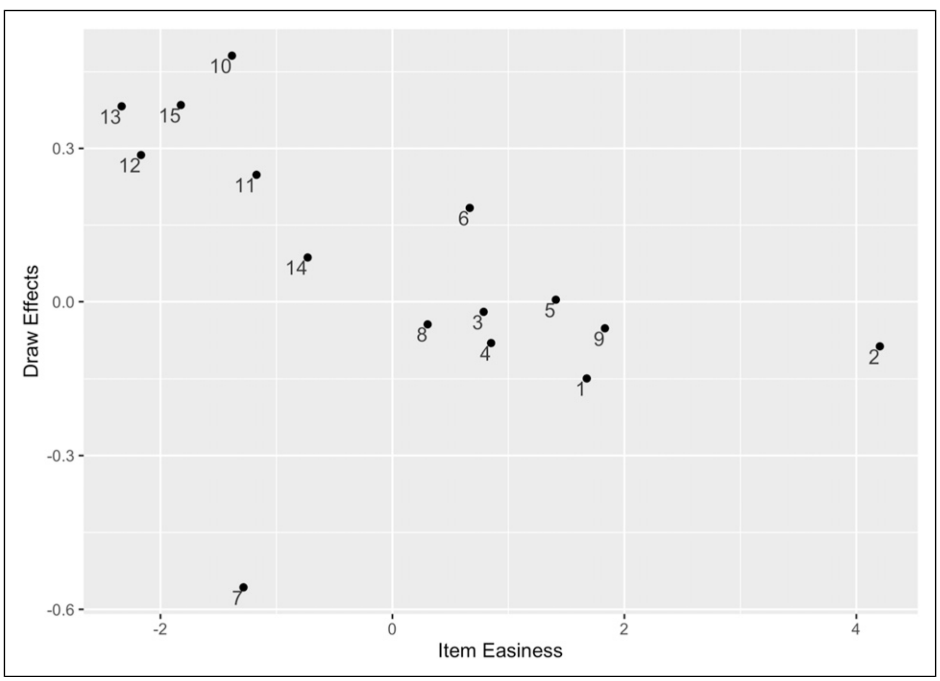

Figure 1. The effect of digital pencil usage by Item difficulty for the whole sample. Source: U.S. Department of Education, Institute of Education Sciences, National center for Education Statistics, National Assessment of Educational Progress (NAEP), response process data from the 2017 NAEP grade 8 mathematics assessment.

between students with LD and their GE peers. Specifically, a significantly lower proportion of students with LD used the digital pencil on more challenging items (Items 5, 7, 10, 11, 12, 13, 14, and 15; item easiness < 0 for all items except 5).

In terms of performance outcomes based on the use of the digital pencil, students with LD who utilized the tool outperformed those who did not on four easier items (Items 3, 5, 6, and 9; item easiness > 0). Conversely, GE students who used the digital pencil generally underperformed compared to those who did not on three of these easier items (Items 1, 4, and 9). Interestingly, for the more difficult Item 7, both students with LD and GE students who used the pencil underperformed compared to those who did not. However, for most difficult items (Items 10, 11, 12, 13, 14, and 15), GE students who used the pencil had better performance outcomes than those who did not—a trend not observed among students with LD.

Model Fit

Our likelihood ratio test, also called the chi-squared test, compares the goodness of fit statistics across three models (Table 3). Model 2 provided better fit than Model 1 ($\chi^2_{df=3}$ = 230.98). Model 3 provided significantly better fit than Model 2 ($\chi^2_{df=9}$ = 1911.62). Model 4 provided significantly better fit than Model 3 ($\chi^2_{df=11}$ = 3599.46). Model 4 was associated with the lowest AIC and BIC, which also suggests best fit among the four models.

Impact of Digital Pencil by Item Difficulty

Our GLMM results did not reveal a significant overall impact of the digital pencil on obtaining correct responses (β_1 = 0.08, p = .26) (Table 4). The significant negative intercept (β_0 = −0.94,

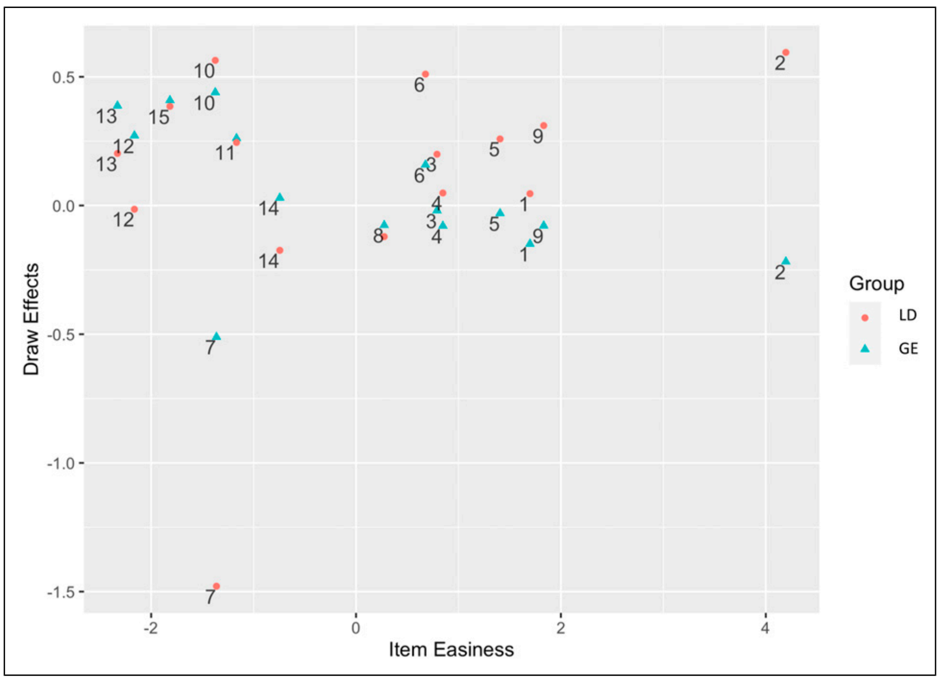

Figure 2. The effect of digital pencil usage by item difficulty for students with learning disabilities versus General education peers. Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), Response Process Data From the 2017 NAEP Grade 8 Mathematics Assessment.

Table 3. Models Evaluation – Goodness of Fit.

Evaluated models	AIC	BIC	Log likelihood	Deviance	Chi-square
Model 1 (1PL IRT)	349605.48	349638.08	−174799.74	349599.48	
Model 2 (add Digital Pencil effect)	349380.50	349445.70	−174684.25	349368.50	230.98*** (df = 3)
Model 3 (add Digital Pencil × LD interaction)	347486.88	347649.88	−173728.44	347456.88	1911.62*** (df = 9)
Model 4 (add Male and Free or Reduced Lunch)	343891.44	344076.11	−171928.71	343857.43	3599.46*** (df = 11)

Note. AIC = Akaike information criterion; BIC = Bayesian information criterion.
***p < .001.
Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress, Response Process Data From the 2017 NAEP Grade 8 Mathematics Assessment.

$p < .001$) indicates a lower logit probability of obtaining a correct answer on an average-difficulty item by an average student without using the digital pencil. Variability in the by-item adjustment to the digital pencil effect ($Var(b_{1i}) = 0.08$) and its negative correlation with item easiness (r = −0.46) suggest that the benefit of the digital pencil is greater on more difficult items for the whole sample.

Figure 1 illustrates the effect size of using the digital pencil across different item difficulties for the whole sample. For easier items (item easiness > 0), there was a negligible or even slightly negative association between digital pencil usage and performance, indicating that digital pencil may not be beneficial for easier items. In contrast, for more difficult items (item easiness < 0), usage of digital pencil was positively associated with problem-solving accuracy.

Impact of Digital Pencil by Item Difficulty and Disability Status

Table 5 reveals that the odds of correctly answering an average-difficulty item correctly are 79% lower for students with LD compared to GE students ($\beta_1 = -1.56$, $p < .001$). There is a negative correlation (r = −0.55) between the by-item adjustment to the digital pencil effect (b_{1i}) and item intercept b_{0i} suggesting that the impact of digital pencil usage decreases when items are easier. Conversely, the correlation between by-item adjustment to

Table 4. GLMM Parameter Estimates for Model 2.

Fixed effect	Estimated effect	s.e.	Z-test	*p*
Intercept	−0.94	0.25	−3.83	<0.001
Digital Pencil	0.08	0.07	1.12	0.26
Random Effect	Variance	SD		
Person (Intercept), b_{0p}	1.51	1.23		
Item (Intercept), b_{0i}	3.47	1.86		
Draw, b_{1i}	0.08	0.28		
Random Effect Correlation Matrix	Draw, b_{1i}			
Item, b_{0i}	−0.46			

Note. s.e. = standard error.
Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress, Response Process Data From the 2017 NAEP Grade 8 Mathematics Assessment.

Table 5. GLMM Parameter Estimates for Model 3.

Fixed effect	Estimated effect	s.e.	z-test	*p*
Intercept	−0.87	0.25	−3.45	<0.001
Digital Pencil	0.06	0.06	1.08	0.28
LD	−1.56	0.16	−9.92	<0.001
Digital Pencil × LD	0.05	0.14	0.36	0.72
Random Effect	Variance	SD		
Person (Intercept), b_{0p}	1.42	1.19		
Item (Intercept), b_{0i}	3.48	1.87		
Digital Pencil, b_{1i}	0.07	0.27		
LD, b_{2i}	0.50	0.71		
Digital Pencil × LD, b_{3i}	0.18	0.42		
Random Effect Correlation Matrix	Item (Intercept), b_{0i}	Draw, b_{1i}	LD, b_{2i}	Draw × LD, b_{3i}
Item (Intercept), b_{0i}	-			
Digital Pencil, b_{1i}	−0.55	-		
LD, b_{2i}	0.21	−0.82	-	
Digital Pencil × LD, b_{3i}	0.75	0.11	−.38	-

Note. LD = students with learning disabilities; s.e. = standard error; SD = standard deviations.
Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress, Response Process Data From the 2017 NAEP Grade 8 Mathematics Assessment.

the LD effect b_{2i} and item intercept b_{0i} is .21, indicating that LD status is more positively associated with problem-solving accuracy on easier items. Moreover, the correlation between the digital pencil-by-LD random effect (b_{3i}) and the item intercept b_{0i} is .75, demonstrating that the benefit of using the digital pencil is more pronounced for students with LD than for GE students on easier items.

Figure 2 displays the relationship between item easiness and the effect size of using digital pencil on correct response probability, differentiated for LD and GE groups (LD represented by circles; GE by triangles). For easy items (easiness >0), the average effect size of using the tool is 0.23 for LD ($\widehat{\beta_{3i}} = 0.23$) but −0.06 for the GE group ($\widehat{\beta_{1i}} = -0.06$) as shown in Table 6. Specifically, for Item 2—the easiest item—the effect size of using pencils on problem-solving accuracy is 0.60 for students with LD ($\beta_{32} = 0.60$) compared with −0.22 for GE students ($\beta_{12} = -0.22$). In contrast, for difficult items (easiness < 0), the effect size of using the tool on problem-solving accuracy is −0.03 for the LD group and 0.18 for the GE group, as per Table 6. Notably, for the most difficult item (Item 13), the effect size was 0.39 for GE students but 0.20 for students with LD.

Table 6. Average Effect Size of Digital Pencil Usage by Groups.

Groups	Average effect size on difficult items (Item easiness <0)	Average effect size on easy items (Item easiness >0)
Whole sample	0.19	−0.03
LD	−0.03	0.23
GE	0.18	−0.06

Note. LD = students with learning disabilities. GE = general education students. Average effect size for the whole sample was estimated from Model 2 (Table 4). Average effect sizes for GE and students with LD were estimated from Model 3 (Table 5).
Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress, Response Process Data From the 2017 NAEP Grade 8 Mathematics Assessment.

Impact of Digital Pencil by Item Difficulty and Disability Status After Controlling for Male and Free or Reduced Lunch Status

The results in Table 7 show that the findings from Model 4 are very similar to results from Model 3 (Table 5), suggesting that the effects identified in Model 3 are robust even when controlling for gender and free or reduced-price lunch status. This indicates that the inclusion of these demographic variables does not significantly alter the relationship between digital pencil usage, item difficulty, and disability status.

Discussion

The digital pencil represents a relatively low-cost and user-friendly technology that can enhance mathematics instruction and learning. This study demonstrates that its benefits extend

Table 7. GLMM Parameter Estimates for Model 4.

Fixed effect	Estimated effect	s.e.	z-test	*p*
Intercept	−0.39	0.34	−1.15	0.252
Digital Pencil	0.07	0.06	1.05	0.29
LD	−1.44	0.18	−8.10	<0.001
Digital Pencil × LD	0.04	0.16	0.24	0.81
Male	−0.13	0.02	−7.77	<0.001
Free or reduced lunch	−0.86	0.02	−52.26	<0.001
Random Effect	Variance	SD		
Person (Intercept), b_{0p}	1.24	1.11		
Item (Intercept), b_{0i}	3.48	1.87		
Digital Pencil, b_{1i}	0.07	0.27		
LD, b_{2i}	0.50	0.71		
Digital Pencil × LD, b_{3i}	0.19	0.43		
Random Effect Correlation Matrix	Item (Intercept), b_{0i}	Draw, b_{1i}	LD, b_{2i}	Draw × LD, b_{3i}
Item (Intercept), b_{0i}	-			
Digital Pencil, b_{1i}	−0.54	-		
LD, b_{2i}	0.21	−0.83	-	
Digital Pencil × LD, b_{3i}	0.71	0.16	−0.42	-

Note. LD = students with learning disabilities; s.e. = standard error; SD = standard deviations.
Source: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress, Response Process Data From the 2017 NAEP Grade 8 Mathematics Assessment.

particularly to accessibility and performance improvements across diverse student populations, including those with LD. By focusing on item-level problem-solving accuracy instead of aggregate test scores, we identified that the effectiveness of the digital pencil is influenced by both the difficulty of the items and the disability status of the students. Our results underscore that the tool is most beneficial when there is an optimal match between task complexity and student ability. Given these findings, it becomes crucial for instructional strategies to be carefully tailored not only to the students' abilities but also to the nature of the task and its complexity.

This research is pioneering in presenting evidence on how students with LD used the digital pencils. Consistent with Yu et al. (2021), who observed underutilization of similar technologies among students with disabilities in digital science notebooks, our findings suggest that students with LD are less likely to utilize digital pencils for more complex tasks, though their usage rates are comparable to their GE peers on simpler tasks. This underutilization could be attributed to the challenging nature of some test items, as indicated by extremely low accuracy rates among students with LD on the most difficult items, suggesting a potential lack of proficiency with using the tools to aid in solving conceptual or multi-step problems. These findings highlight the importance of explicit instruction to support students in making effective use of digital tools like the digital pencil, especially for those who may struggle with complex tasks.

We further explored the relationship between different types of problems (e.g., word problems, calculations, conceptual tasks) and digital pencil usage. Our analysis revealed that students with LD benefitted more from using the digital pencil on simpler, procedural tasks (e.g., basic calculations or measurements). Specifically, students with LD who used the digital pencil outperformed their peers who did not on Items 3, 5, 6, and 9—items that involved straightforward calculations or concrete procedures. This suggests that the digital pencil helps students with LD externalize their thinking and better manage cognitive load on tasks that involve familiar steps or require organization of calculations.

In contrast, for more complex, conceptual items (e.g., Items 10, 11, 14, 15), GE students benefitted more from using the digital pencil. These items required abstract reasoning, multi-step processes, or geometric and algebraic thinking, where the ability to visualize relationships and organize multi-step processes played a crucial role. The digital pencil appears to support GE students in externalizing complex thought processes, allowing them to organize information more effectively, particularly for challenging tasks that required higher-order reasoning. This suggests that the pencil's utility may differ depending on the type of problem: students with LD benefit more on simpler, procedural tasks, while GE students are able to leverage the pencil for more abstract, complex problems.

While prior studies have established the benefits of drawing mathematical representations in bolstering mathematics performance (van Garderen et al., 2012; Garderen et al., 2014), they have not dissected how these advantages vary with task type, task difficulty, and disability status. Our work addresses this void, illustrating that digital pencils can improve outcomes for GE students on challenging tasks and students with LD on less demanding ones, enhancing odds of correct answers by 20% and 26%, respectively. This differential impact reinforces the need for tailored instructional strategies that take into account both the nature of the task and student ability.

For instance, the benefit of the digital pencil may be tied to the type of cognitive processes the task requires. For calculation-heavy tasks, students with LD may find the pencil helpful in externalizing and organizing their thoughts, enabling them to focus on step-by-step problem-solving. In contrast, for word problems or conceptual items, which require more complex cognitive engagement (e.g., interpretation, reasoning), the pencil's utility may be less about solving the problem directly and more about managing cognitive load. For GE students, the pencil seems to facilitate higher performance on more complex and multi-step items, suggesting that when paired with advanced problem-solving skills, the tool can help them better visualize and organize their work, especially for challenging conceptual or word-based tasks.

The crux of this discovery is that the efficacy of digital pencils hinges on an optimal match between task complexity, type, and student ability—emphasizing these tools' role as cognitive offloading mechanisms. By facilitating the externalization of cognitive processes, digital pencils minimize cognitive overload, aiding students in navigating complex mathematical terrains. Yet, this advantage reaches its zenith only when the tasks align with the learners' current proficiency levels. This also implies that the type of problem—whether it involves multi-step reasoning, conceptual understanding, or basic calculation—can further mediate how the tool is used. In this way, the tool can function as a scaffold, allowing students with LD to engage more deeply with challenging material, although its immediate impact may be more visible on simpler tasks.

Given the importance of matching student ability with both task type and difficulty, educators should carefully consider how digital pencils are integrated into instruction. Teachers could implement targeted strategies such as explicit instruction on how and when to use this tool. While the literature on teaching students to use digital pencils specifically is limited, there is substantial research on the broader topic of teaching students to use visual representations in problem-solving. For example, research on visual representations in problem-solving suggests that students benefit from explicit instruction in using tools like diagrams, drawings, and other visual aids to externalize their cognitive processes, leading to significantly improve students' understanding of complex concepts (Schwamborn et al., 2010; van Meter & Firetto, 2013; van Meter & Garner, 2005). In a digital learning context, Xie and Zhou (2024) emphasize the importance of clear, structured guidance for students using digital tools like touchscreen interfaces. These insights suggest that teaching students how to effectively use digital pencils could involve similar strategies, such as modeling tool usage, providing guided practice, and offering feedback on their application.

1. Modeling: Demonstrate how to use the digital pencil to solve various types of problems, such as drawing diagrams, number lines, and geometric shapes, plotting graphs, or organizing multi-step calculations.
2. Guided practice: Allow students to practice using the digital pencil with teacher guidance, offering prompts or scaffolding to support their understanding of how to apply the tool to different problems.
3. Feedback: Offer targeted feedback on how students use the digital pencil, emphasizing strategies for improving accuracy and efficiency when solving problems.

To integrate digital pencils effectively into mathematics instruction, teachers should consider the following strategies to tailor to different ability levels, task types, and task difficulties. For students with LD, digital pencils appear to be most beneficial for simpler calculation-based tasks, where the tool helps externalize thinking and reduce cognitive load. Teachers should initially introduce the digital pencil in tasks that match the student current proficiency level where students feel confident and then gradually incorporate its use into more complex tasks. As students gain proficiency, teachers can introduce more complex problems that encourage the use of the digital pencil for step-by-step problem solving. For GE students, the digital pencil is particularly helpful for more difficult and conceptual tasks, making it a great tool for supporting the organization of information, visualization of abstract concepts, and exploration of multiple solution strategies.

Professional development is essential for ensuring that teachers are equipped to integrate these tools effectively. Professional development should emphasize not only the technical aspects of integrating digital tools but also the need to differentiate instruction based on the complexity of tasks, task types, and student abilities, ensuring that both students with LD and GE students maximize the benefits of digital pencils. Such training could cover the pedagogical strategies for using digital pencils to support problem-solving, as well as techniques for differentiating instruction based on student ability and task difficulty. By equipping teachers with these skills, schools can enhance the overall effectiveness of digital tools in the classroom.

Despite these insights, several limitations must be considered when interpreting these findings. First, this study did not randomly assign students to either use or not use the digital pencil for solving mathematics problems. It is possible that students who used the digital pencil may already possess stronger problem-solving skills. Future studies should employ randomized controlled trials, stratified by students' familiarity and training in using digital tools, to rigorously investigate the interaction between task type, problem difficulty, and digital pencil usage.

Second, the process data do not distinguish between the use of a stylus or a finger when students used the embedded pencil tool. Future research should examine whether different input methods (e.g., stylus or finger) influence students' problem-solving accuracy. Additionally, the NAEP process data do not capture the product of students' drawings, making it impossible to evaluate the type and accuracy of their drawings (e.g., visual-schematic or pictorial representations). Future studies should aim to collect more granular process data, such as screen recordings and detailed logs of student interactions with digital tools, to better understand how students use these tools—whether for drawing, calculation, or other purposes—and how this usage relates to their academic performance.

Third, some students who chose not to draw might have generated images internally through imagination (Cheng & Beal, 2020). This study did not differentiate between internal visualization and no visualization at all.

Fourth, future research should explore effective instructional methods for teaching students to use digital pencils, particularly in mathematics education. Experimental studies comparing various teaching approaches—such as direct instruction, guided practice, or peer learning—could offer valuable insights into the best practices for integrating digital tools into the classroom and supporting students in maximizing their potential.

Fifth, this study used pre-pandemic data, and future research should replicate the analysis using more recent NAEP data to account for changes in instruction and learning resulting from the pandemic. Additionally, future research should examine demographic variables, such as socioeconomic status, locale, gender, and ethnicity, to better understand disparities in digital tool usage. Expanding the study to include 4th-grade students could also provide further insights into how digital pencil usage impacts students across different age groups. By addressing these limitations and continuing to refine our understanding of how digital pencils can be effectively integrated into educational practices, we can better tailor intervention programs to individual students' abilities.

Conclusion

This study offers valuable insights into how digital pencil usage is associated with mathematics performance among students with LD and their GE peers. Our findings show that the effectiveness of digital pencils is closely linked to both student ability and task difficulty—with the tool being most beneficial when there is an optimal match between these factors. Specifically, digital pencils improve problem-solving accuracy for GE students on more challenging tasks, while students with LD benefit most from the tool on simpler tasks. This underscores the need for tailored instructional strategies to help students make the most of these digital tools based on their individual skill levels and the complexity of the tasks they face.

Additionally, the study highlights the importance of explicit instruction in how to use digital tools like the digital pencil. Teachers should incorporate guided practice, feedback, and modeling into their lessons to ensure that all students—especially those with disabilities—can effectively engage with digital tools. By integrating these strategies and ensuring equitable access to training and digital tools, educators can help bridge the digital divide and ensure that all students, regardless of background or ability, benefit from these technologies.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R324P230002 to Digital Promise. The opinions expressed are those of the authors and do not represent the views of the Institute or the U.S. Department of Education.

ORCID iD

Xin Wei https://orcid.org/0000-0002-6978-6609

Note

1. The minimum detectable effect size (MDE) for comparing math performance between students with LD and GE peers is 0.07, based on a two-sided 5% significance level. This calculation assumes that 20% of the variance in outcomes can be explained by student background characteristics.

References

Boeck, P. D., Bakker, M., Zwitser, R., Nivard, M., Hofman, A., Tuerlinckx, F., & Partchev, I. (2011). The estimation of item response models with the lmer function from the lme4 package in R. *Journal of Statistical Software*, *39*(12), 1–28. https://doi.org/10.18637/jss.v039.i12

Boonen, A. J. H., van Wesel, F., Jolles, J., & van der Schoot, M. (2014). The role of visual representation type, spatial ability, and reading comprehension in word problem solving: An item-level analysis in elementary school children. *International Journal of Educational Research*, *68*(4), 15–26. https://doi.org/10.1016/j.ijer.2014.08.001

Booth, J. G., & Hobert, J. P. (1999). Maximizing generalized linear mixed model likelihoods with an automated monte carlo EM algorithm. *Journal of the Royal Statistical Society: Series B*, *61*(1), 265–285. https://doi.org/10.1111/1467-9868.00176

Cheng, L., & Beal, C. R. (2020). Effects of student-generated drawing and imagination on science text reading in a computer-based learning environment. *Educational Technology Research & Development*, *68*(1), 225–247. https://doi.org/10.1007/s11423-019-09684-1

Garderen, D. V., Scheuermann, A., & Poch, A. (2014). Challenges students identified with a learning disability and as high-achieving experience when using diagrams as a visualization tool to solve mathematics word problems. *ZDM: International Journal on Math Education*, *46*(1), 135–149. https://doi.org/10.1007/s11858-013-0519-1

Geary, D. C. (2004). Mathematics and learning disabilities. *Journal of Learning Disabilities*, *37*(1), 4–15. https://doi.org/10.1177/00222194040370010201

Griffin, S. A., Case, R., & Siegler, R. S. (1994). Rightstart: Providing the central conceptual prerequisites for first formal learning of arithmetic to students at risk for school failure. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 24–49). MIT Press.

Grigorenko, E. L., Compton, D. L., Fuchs, L. S., Wagner, R. K., Willcutt, E. G., & Fletcher, J. M. (2020). Understanding, educating, and supporting children with specific learning disabilities: 50 years of science and practice. *American Psychologist*, *75*(1), 37–51. https://doi.org/10.1037/amp0000452

Krawec, J., Huang, J., Montague, M., Kressler, B., & Melia de Alba, A. (2012). The effects of cognitive strategy instruction on knowledge of math problem-solving processes of middle school students with learning disabilities. *Learning Disability Quarterly*, *36*(2), 80–92. https://doi.org/10.1177/0731948712463368

Krawec, J. L. (2014). Problem representation and mathematical problem solving of students of varying math ability. *Journal of Learning Disabilities*, *47*(2), 103–115. https://doi.org/10.1177/0022219412436976

Kwak, M., & Gweon, G. (2019). Should students use digital scratchpads? Impact of using a digital assistive tool on arithmetic problem-solving. In S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, & R. Luckin (Eds.), *Artificial intelligence in education. AIED 2019 Lecture Notes in Computer Science* (Vol. 11626, pp. 153–157). Springer. https://doi.org/10.1007/978-3-030-23207-8_29

Lee, I. C., & Cheng, P.-J. (2021). The influence of different drawing tools on the learning motivation and color cognition of the fourth grade students at the elementary school. Paper presented at the HCI international 2021 - posters, Cham.

Mayer, R. E. (1985). *Mathematical ability. Human abilities: An information processing approach* (pp. 127–150). Freeman.

Moscatelli, A., Mezzetti, M., & Lacquaniti, F. (2012). Modeling psychophysical data at the population-level: The generalized linear mixed model. *Journal of Vision*, *12*(11), 26. https://doi.org/10.1167/12.11.26

National Assessment of Educational Progress. (2020). NAEP students with disabilities questionnaire. https://nces.ed.gov/nationsreportcard/subject/experience/pdfs/sd_2020.pdf (Accessed 26 June 2021).

National Assessment of Educational Progress. (2022). Reading and mathematics scores decline during COVID-19 pandemic. https://www.nationsreportcard.gov/highlights/ltt/2022/

National Center for Education Statistics. (2024). Students with disabilities. In *Condition of education*. U.S. Department of Education, Institute of Education Sciences. Retrieved Oct 1, 2024, from. https://nces.ed.gov/programs/coe/indicator/cgg

National Council of Teachers of Math (NCTM). (2000). *Principles and standards for school mathematics*. NCTM.

National Research Council. (2001). *Adding It Up: Helping Children Learn Mathematics*. Washington, DC: The National Academies Press. https://doi.org/10.17226/9822

Patti, A. L., & Garland, K. V. (2015). Smartpen applications for meeting the needs of students with learning disabilities in inclusive classrooms. *Journal of Special Education Technology*, *30*(4), 238–244. https://doi.org/10.1177/0162643415623025

Quillin, K., & Thomas, S. (2015). Drawing-to-learn: A framework for using drawings to promote model-based reasoning in biology. *CBE-Life Sciences Education*, *14*(1), es2–16. https://doi.org/10.1187/cbe.14-08-0128

Rau, M. A. (2017). Conditions for the effectiveness of multiple visual representations in enhancing STEM learning. *Educational Psychology Review*, *29*(4), 717–761. https://doi.org/10.1007/s10648-016-9365-3

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rose, D. H., Meyer, A., Strangman, N., & Rappolt, G. (2016). *Teaching every student in the digital age: Universal design for learning*. Association for Supervision and Curriculum Development.

Rubin, A., Storeygard, J., & Koile, K. (2015). Supporting special needs students in drawing mathematical representations. In T. Hammond, S. Valentine, A. Adler, & M. Payton (Eds.), *The impact of pen and touch technology on education. Human–Computer Interaction Series*. Springer. https://doi.org/10.1007/978-3-319-15594-4_6

Schwamborn, A., Mayer, R. E., Thillmann, H., Leopold, C., & Leutner, D. (2010). Drawing as a generative activity and drawing as a prognostic activity. *Journal of Educational Psychology*, *102*(4), 872–879. https://doi.org/10.1037/a0019640

Stijnen, T., Hamza, T. H., & Özdemir, P. (2010). Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. *Statistics in Medicine*, *29*(29), 3046–3067. https://doi.org/10.1002/sim.4040

van Garderen, D., Scheuermann, A., & Jackson, C. (2012). Developing representational ability in math for students with learning disabilities: A content analysis of sixth and seventh grade textbooks. *Learning Disability Quarterly*, *35*(1), 24–38. https://doi.org/10.1177/0731948711429726

van Meter, P., & Firetto, C. M. (2013). Cognitive model of drawing construction: Learning through the construction of drawings. In G. Schraw, M. T. McCrudden, & D. Robinson (Eds.), *Learning through visual displays* (pp. 247–280). Information Age Publishing Inc.

van Meter, P., & Garner, J. (2005). The promise and practice of learner generated drawing: Literature review and synthesis. *Educational Psychology Review*, *17*(4), 285–325. https://doi.org/10.1007/s10648-005-8136-3

Vrieze, S. I. (2012). Model selection and psychological theory: A discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods*, *17*(2), 228–243. https://doi.org/10.1037/a0027127

Wammes, J. D., Roberts, B. R. T., & Fernandes, M. A. (2018). Task preparation as a mnemonic: The benefits of drawing (and not drawing). *Psychonomic Bulletin & Review*, *25*(6), 2365–2372. https://doi.org/10.3758/s13423-018-1477-y

Wang, T., Graves, B., Rosseel, Y., & Merkle, E. C. (2022). Computation and application of generalized linear mixed model derivatives using lme4. *Psychometrika*, *87*(3), 1173–1193. https://doi.org/10.1007/s11336-022-09840-2

Way, D., & Strain-Seymour, E. (2021). A framework for considering device and interface features that may affect student performance on the National Assessment of Educational Progress. Retrieved on Nov 1 2023 from. https://www.air.org/sites/default/files/Framework-for-Considering-Device-and-Interface-Features-NAEP-NVS-Panel-March-2021.pdf

Xie, H., & Zhou, Z. (2024). Finger versus pencil: An eye tracking study of learning by drawing on touchscreens. *Journal of Computer Assisted Learning*, *40*(1), 49–64. https://doi.org/10.1111/jcal.12863

Yu, J., Wei, X., Hall, T. E., Oehlkers, A., Ferguson, K., Robinson, K. H., & Blackorby, J. (2021). Findings from a two-year effectiveness trial of the science notebook in a universal design for learning environment. *Frontiers in Education*, *6*(6), 437. https://doi.org/10.3389/feduc.2021.719672

Author Biographies

Xin Wei possesses extensive expertise in statistical modeling, measurement, research design, and special education, and is dedicated to leveraging these skills to enhance educational outcomes for students with disabilities.

Susu Zhang is an Assistant Professor of Psychology and Statistics at the University of Illinois Urbana-Champaign. Her research involves latent variable modeling for testing and learning data, including the analysis of complex data (e.g., log data) to address measurement and educational questions.

Jihong Zhang is an Assistant Professor of Educational Statistics and Research Methods at the University of Arkansas, specializing in psychometrics and advanced statistical modeling, with research interests in psychological network analysis, Bayesian latent variable modeling, and Item Response Theory.