Narrowing the Digital Divide in Early Maths: How Different Modes of Assessment Influence Young Children’s Mathematical Test Scores

Sian McDonald & Olga Fotakopoulou

To cite this article: Sian McDonald & Olga Fotakopoulou (28 Sep 2023): Narrowing the Digital Divide in Early Maths: How Different Modes of Assessment Influence Young Children's Mathematical Test Scores, Early Education and Development, DOI: 10.1080/10409289.2023.2260683

To link to this article: https://doi.org/10.1080/10409289.2023.2260683

© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC.
Narrowing the Digital Divide in Early Maths: How Different Modes of Assessment Influence Young Children’s Mathematical Test Scores

Sian McDonalda and Olga Fotakopouloub

aNHS, Youth Psychological STR Worker; bDevelopmental Psychology, Birmingham City University

ABSTRACT

Research Findings: Children are increasingly using touchscreen technologies at home, which has become a recurring feature within their classroom too. Research has investigated the potential effect of using computer-based tests to assess pupils’ performance rather than traditional paper tests. An agreement has still not been formed about the impact of the mode of assessment on pupils’ mathematical test scores. A mixed methods design was employed to explore the impact of the testing on young children’s mathematical test scores. Thirty-seven children 4-7-years old were recruited from a primary school in England with their parents. A mathematical test on paper and an iPad was administered to each participant which was accompanied by an interview. Data regarding the use of touchscreens at home were explored with an on-line parental questionnaire. The results showed that gender and test type impact mathematical test scores, with females performing the best on the iPad test. The findings also suggested that as tablet usage increased at home, iPad mathematical test scores decreased. The interviews revealed that children rely on different strategies when resolving mathematical problems.

Practice or Policy: Digital testing may enable a better investigation of mathematical skills in the first years of schooling and of differences between males and females’ responses to solving mathematical questions, which then could be used to tailor the curriculum.

Technological advancements of the 21st century have created a remarkable shift in learning experiences for young children (Kim et al., 2021; Kucirkova et al., 2019; Schmid et al., 2023). The usage of technology has significantly increased for children between the ages of 0–8 years old over the last decade (Hadlington et al., 2019). Children as young as two-years old are having daily screen time (ST) on tablets and different devices (Chenchen & Gwo-Jen, 2023). This high exposure has turned the 21st century children to digital natives who naturally use technologies as the components of their everyday lives in both formal and informal settings (Li et al., 2023). The World Health Organisation have insisted that sedentary ST should be no more than one hour in 24 hours for children under the age of 5-years old (World Health Organisation, 2019). Still, there is conflicting evidence for children’s ST and whether tablets impact learning (Kalati et al., 2022; Schmid et al., 2023). The overwhelming popularity of tablets has contributed to an educational uptake, which has positively supported many subjects (González-González et al., 2019). Herodotou (2018) reviewed 19 studies examining the touchscreen learning effects on young children and reported generally positive effects on literacy, mathematics, science learning, problem-solving skills and self-efficacy. Though, there were no studies reporting the

Tablets introduce challenges to the cognitive load of children; they have been found to offer a highly interactive learning experience (McEwen & Dubé, 2015). As acknowledged by OECD (2019), “active interaction with touchscreens can generate dynamic stimulation, and, if used appropriately, may be as engaging and cognitively stimulating as traditional toys or books” (p. 45). Xie et al. (2018) contacted a meta-analysis for the learning achievements of touchscreen mobile devices for young children’s learning; they examined empirical articles about the overall effects of touchscreen devices on young children’s learning performance as well as moderators of this effect and their findings suggest that touchscreen technology usage from a young age has a positive relationship with school achievement. Research investigating the impact of tablet use on children’s mathematical skills demonstrate the improvement of problem-solving and mathematical ability overall (Papadakis et al., 2016). Schacter and Jo (2017) investigated mathematics performance of 273 young children after attending tablet-based math instructions for 15 weeks and found that children demonstrated 12 times higher gain in their mathematical skills. Disney et al. (2019) investigated the impact of iPad-based games on numeracy learning in 4-year-olds; the researchers used games on the gestural interface devices to explore learning of numeracy concepts. They found that young children’s numeracy learning outcomes were improved (Disney et al., 2019). Shanley et al. (2020) found improved accuracy in responding to mathematical problems in a tabled-based intervention for 5-to 6-year-old children with mathematical difficulties. The above-mentioned study investigated functional relations between the provision of instructional cueing and self-regulation support features in an iPad-based mathematics intervention program and improved response accuracy for students.

Moreover, the potentials of touchscreens to facilitate personalized, flexible, and mobile learning experiences, as well as individualized assessment and rich communication (Kalati et al., 2022), have led to the introduction of these devices into school settings for multiple educational purposes (Lee, 2023; McLean, 2016). Yet, despite the growing body of research on the relationship between tablet use and student learning, there is a dearth of research comparing the effects of iPad technology and paper-and-pencil-delivered interventions on student mathematic outcomes (Kromminga & Codding, 2021). There are limited empirical studies that investigate the suitability of integrating digital technology in the assessment of learning of mathematical concepts. Furthermore, there is scarce research examining whether using tablets to administer a mathematical test has an impact on children’s ability to solve mathematical problems compared to traditional methods.

**Rationale**

Numerous studies have focused on the impact technology has on the learning experience. In the modern educational context, emphasis is placed on educating pupils/learners in a way that they can monitor their learning, develop their knowledge and skills and assessing their results (Vasileiadou & Karadimitriou, 2021). In recent decades, technology has influenced various aspects of assessment in mathematics education; though, it is at risk of focusing too much on assessment of lower order goals, such as the reproduction of procedural, calculation based, knowledge and skills (Hoogland & Tout, 2018). Standardised testing is prevalent at many levels of schooling (Csapó & Molnár, 2019), although teachers’ conceptions largely determine whether and how these instruments are used (Frans et al, 2020). But what is the impact of testing practices in schools, especially technology-based testing practices and how can the use of technology/iPads impact young children’s mathematical skills? And what the implications might be for schools that embrace the use of technology/iPads for the assessment of their learners?

Recent research has suggested that females use technology more for educational purposes, as opposed to males participating in consumptive activities (Johansen, 2019; Marsh et al., 2018). Watson et al. (2023) with their exploratory investigation into the factors related to the use of technology among girls in Kenya found that use of technology could especially benefit girls, who may have higher engagement than boys when afforded the same level of technology access. Therefore, technology within the classroom could be advantageous for females’ performance. And how can
Technology-based testing practices influence academic attainment in girls and boys at this crucial stage of their mathematical understanding and development of mathematical reasoning.

**Tablet Technology Use with Young Children**

Technology usage has significantly increased for children between the ages of 0–8 years old (Dardanou et al., 2020; Koran et al., 2022; Kumar & Mohite, 2018; O’Connor et al., 2023). Children between the ages of 4–6 years old predominantly use tablets for gaming and watching cartoons, with educational apps being the last (Nikolopoulou, 2020). O’com (2019) reported that children between the ages of 3–4 years old are using their tablets often. However, this is not always concerning, as Marsh et al. (2018) suggested that children under 5-years old engaged with play-based apps can develop their creativity skills. But it was not clarified how this can translate to academic skills.

Gender differences occur within the world of technology, as technology has been associated with being male territory (Mims-Word, 2012). Regarding gender and roles, authors such as Axell and Boström (2021) pointed out that technology is strongly connected to the female/male dichotomy, reporting gender differences in the attitudes toward technologies in terms of education. Although males have a higher usage and increased positive attitude toward technology (Romero-Tena et al., 2022), there are no gender differences of ability of using digital technologies (McKenney & Voogt, 2010). However, females have been found to have a higher visual learning style and use educational apps more frequently than males (Nang et al., 2015; Pruet et al., 2016), resulting in higher benefits for female technology users. Young children are often drawn toward the fun aspect of tablets rather than educational purposes. Research has found that although educational apps are less likely to be used, females use them more frequently (Nang et al., 2015). These results could be occurring due to the apps the different genders use on their tablets. Therefore, it is essential to assess both genders when investigating the impact of technology on different learning activities and assessment of educational attainment.

**Young Children’s Learning Through Tablet Technology**

Research suggests that touchscreen technology could potentially be a profound learning tool for young children (Kewalramani et al., 2020; Oliemat et al., 2018). Evidence suggests that tablets enhance children’s emerging literacy skills (Neumann & Neumann, 2014; 2017), science concepts (Furman et al., 2019), problem solving, and mathematical skills (Herodotou, 2018). For example, tablet usage in science has increased students’ interest within the subject, as lessons are made more interactive with visual aids (Kewalramani et al., 2020). Research investigating the impact of tablet use on children’s mathematical skills, principally describe studies involving an intervention. Papadakis et al. (2016) investigated and compared the influence of using computers and tablets in the development of mathematical competence in early childhood education. They implemented a 14-week intervention, where children were taught mathematics with the use of the same educational software on computers and laptops. The results showed that teaching with the use of tablets has contributed significantly to the improvement of children’s mathematical abilities. Volk et al. (2017) examined the impact of tablet-based, cross-curricular math activities on the acquisition of higher-order learning outcomes over seven months in twelve third grade classrooms in Slovenia. The authors concluded that in cross-curricular math teaching, tablets offer efficient use of resources from different subjects and multiple representations which facilitate learning outcomes in the cognitive, affective-social and psychomotor learning domains. Such interventions have demonstrated the improvement of problem-solving and mathematical ability and these findings were applicable to many age groups, but how children’s usage and performance varied between genders was not explored (Papadakis et al., 2016; Volk et al., 2017).

Outhwaite et al. (2017) concluded that when tablet software is age appropriate and grounded in a well-designed curriculum, then it results in being an effective support for early math development. The researchers used a tablet intervention and findings of mathematical
improvement for low-achieving children between the ages of 4–7-years old were significant. Interventions like the one mentioned above can be highly beneficial in closing the gap in early math attainment from the start of primary school. Interventions such as “Math Shelf” have exhibited a significant improvement on mathematical achievement (Schacter & Jo, 2017). Play-based learning with tablet apps has also been explored. Miller (2018), for example, introduced an interactive mathematical app to young children using an iPad. Only a small improvement in mathematical achievement was noted but the interactive game encouraged group-work. However, according to Miller (2018), factors influencing the use of interactive technology included the quality of the app such that creative and fun apps promoted children’s engagement in learning mathematics. The level of difficulty of an app was a second factor influencing children’s use of interactive technology. If the difficulty level was too challenging, children became disengaged with the app. Kosko and Ferdig (2016) reported that well-designed mathematics apps improved achievement and concluded that well-designed mathematics apps can support student learning, but more research is needed to explore the extent to which these apps improved learning, and this is reflected in children’s mathematical test scores.

Previous literature has demonstrated that using technology as a mathematical intervention as well as play-based learning has been seen to improve young children’s mathematical skills (Miller, 2018; Schacter & Jo, 2017). Yet research that focuses on the ways technology affects the learning experience does not explore technology as a test type nor whether test scores on a digital device differ when compared to paper-based test scores. This type of research also lacks populations under the age of 7-years-old. Our research aimed to fill this gap by examining mathematical test scores and using different modes of assessment in different year groups of primary school children in England.

This study utilized a paper based, and an iPad based mathematical test, which involve the same tasks but have different answers to explore whether and to what degree there were differences in test scores between the paper and iPad test. This would reveal if there were any overall differences in test scores between the paper and iPad test. The results would indicate if any immediate effects of tablet technology on mathematical test scores occurred when an intervention has not been implemented. Further examination explored whether there were differences between the age and gender of students on test scores. And the addition of semi-structured interviews of children after the tests were utilized to determine whether children had different problem-solving techniques/strategies when answering test questions.

**Research Questions and Hypotheses**

This present study adopted a mixed methods approach to investigate if technology has an impact on mathematical test scores and why the result has occurred. It explored whether and how mathematical testing influences how young children answer questions posed and whether factors such as year group and gender influence their test-answering performance on counting, knowledge of shapes, addition and subtraction.

The following research questions were explored:

- Does age, gender and test type (paper/iPad) influence mathematical test scores?
- How does children’s mathematical technique change when a test is on paper compared to an iPad?

The following hypotheses were established:

- If a child uses an iPad/tablet device frequently, then their test scores will be higher for the iPad test.
- Children of both genders will use different strategies when answering/resolving the mathematical problems.
Methods

Design

A mixed-methods approach was employed, utilizing a counterbalanced within-subjects design. The dependent variables are test scores, technology usage at home and strategy of how to answer mathematical questions; the independent variables are gender, year group and test type (paper or iPad).

Participants

In total, 37 students (17 female and 20 male) between the ages of 4–7 years old participated. Participants were split into year groups: Reception, Year 1 and Year 2 (see Table 1). Inclusion criteria insisted that participants must speak fluent English and attend a primary school within the United Kingdom. Exclusion criteria insisted that participants must not have a diagnosed learning difficulty and/or a difficulty with mathematics.

Research Materials (Please Find Them on https://osf.io/yqfm5/)

Experimental Tasks

An iPad was used to present the iPad version of the mathematical test, which was generated using Qualtrics software, Version September 2019 of Qualtrics.

Mathematics Test

The experimental task measures mathematical ability for addition, subtraction, counting and shapes (see Appendices A-F on Open Science Framework). These tests are self-constructed, based on Piaget’s theory of cognitive development (Ojose, 2008) and reviewing the school curriculum and online platforms (IXL- https://uk.ixl.com/math/). There were six variations of the test: three paper tests and three iPad tests. Each year group had their own paper and iPad test tailored to the mathematical ability of that year group. The students in reception had images to aid in the understanding of their mathematical questions, whereas Year 1 students had some questions with images and some without (see Appendices A-F). A pilot study was conducted prior to the main data collection phase to ensure the suitability of the items. After completing the pilot study, the internal reliability of the mathematical test assessing the consistency of results across items within the test was \( r = .75 \) and was considered satisfactory. Also, the mathematical test measured the concepts/skills it was designed to evaluate as it corresponded accurately to pupils’ scores in Maths according to their teachers.

Parental Questionnaire

A self-constructed on-line questionnaire (see Appendix G) for the parents was used and adapted from O’Connor and Fotakopoulou’s (2016) study. It measured how much time a participant spent using technology and what they spent their time doing. As well as asking for demographic information about the participant. There are eleven questions and duration was approximately five minutes.

| Table 1. Demographic information of participants including the total, year group and gender. |
|-----------------|-----------------|-----------------|-----------------|
| Year Group      | \( N \)          | Female \( n \)  | Male \( n \)    |
| Reception       | 11              | 6               | 5               |
| Year 1          | 14              | 6               | 8               |
| Year 2          | 12              | 5               | 7               |
| Total           | 37              | 17              | 20              |
Semi-Structured Interview Schedule

The interview questions (see Appendix H) were part of a metacognitive discourse from Shilo and Kramarski (2018) study of metacognitive mathematical discourse in the classroom. These questions were chosen as research has shown that using metacognitive mathematical discourse improves metacognition and problem-solving gains (Dignath & Büttner, 2008). The semi-structured interview with the children from each year group lasted approximately 10 minutes and explored the techniques/strategies used and reflection on the whole process of resolving a problem.

Children were asked overall eight questions within the semi-structured interview; two questions exploring planning (e.g. Did you know how to solve this question?), three questions exploring monitoring (e.g. Where have you seen these types of questions before?) and three questions examining their ability to reflect on the process of resolving the mathematical problems (e.g. Did you feel like you managed to answer the question well? Why?). Planning involves the selection of appropriate strategies and the allocation of resources. Monitoring refers to checking one’s comprehension and performance, e.g., by means of self-testing. Metacognitive reflection referred to their ability to reflect on the metacognitive strategies used (Dignath & Büttner, 2008).

Questions were designed in a way that would facilitate an open conversation between participant and researcher. The questions were asked in the best possible way the child would understand; therefore, some terminology and words had been amended to be suitable for these age groups. As we are aware of the potential biases in children’s responses during the interviews with them, at the planning phase of our research, we paid particular attention to the way the interview was set up and the questions were developed. We insisted on the open nature of our questions in order to elicit genuine answers and, as we are aware of the confirmation bias, we were not looking for specific answers. We made sure that we were reading the question the same way for all participants, not prompting for answers, re-reading questions exactly as they are. Children were not interviewed in a room with mathematical displays on the wall. Also, children were not interviewed after a math lesson had taken place. As the research team were aware that children tend to comply with adults and “authority figures” ensured that we were open to their responses and to whatever they had to say. We invested in building rapport with our interviewees and drew upon our experience with working with children in different educational settings. As we have been extremely careful with our interviewing style, we made sure that in our answers, we didn’t overly confirm their responses, demonstrating a neutrality which alongside the open nature of the questions could minimize any potential biases. We applied the same approach to the way we handled the data also; all interviews were recorded and during analysis, we looked for any indicators that biases were involved in children’s responses and didn’t detect any indicators of biases. We are aware that all research could involve some biases; those carrying out qualitative research are an integral part of the process and final product, and separation from this is neither possible nor desirable. The concern instead should be whether we have been transparent and reflexive (i.e., critically self-reflective about our own preconceptions, relationship dynamics, and analytic focus (McConaughy & Whitcomb, 2022) about the processes by which data have been collected, analyzed, and presented. The interviews were transcribed by the research team.

Procedure

Parents who consented to their child’s participation were sent the parental questionnaire. Prior to main data collection a pilot study was conducted with a child from each year group to assess all materials were appropriate for their age. After the completion of the pilot study, all necessary changes were implemented to the research tools and researchers ensured that reliability and validity were met. Each participant was given an assent form and the researcher read it to the participant and asked for their consent. Half of the participants completed the mathematical test on paper first and the other half completed the Qualtrics test first. The researcher verbally read out each question to the participants and wrote the answer if the participant was unable to. The participants who completed the paper test were administered the semi-structured interview questions. The interview was recorded on
a Dictaphone, to refer back to for analysis. The participant was then thanked and given a sticker. A week’s interval was then taken. After this, the researcher returned to the school and participants completed the other test. Paper participants then took part in the interview and all participants were thanked and given a sticker.

Personal data such as the consent forms and raw data were kept on my password protected laptop on One Drive. All data were stored in accordance with the university’s data protection policy. The interviews were transcribed, anonymized and any identifying information was removed. The recordings of the interviews were kept until the analysis was completed, after which it was destroyed.

**Analysis**

In our analysis, we included the complete protocols collected from 37 participants. Protocols which were not completed were not taken into consideration.

**Mathematics Tests**

A three-way mixed analysis of variance (ANOVA) was computed with gender and year group as between-subject factors and test type as a within-subject factor. Bonferroni Correction was used as the post hoc test where appropriate. Kruskal-Wallis H test was conducted with all year groups and both iPad and paper test sub-scores. This test was adopted, as data violated the assumption of being normally distributed. Bonferroni Correction was used again for post hoc. A Mann-Whitney U test was conducted with gender and both iPad and paper test scores. This test was adopted as data was not normally distributed.

**Parental Questionnaire.** A Spearman’s Rank-Order Correlation was run to assess the relationship between average time spent on iPad at home and iPad mathematical test scores on the different tasks.

**Semi-Structured Interviews.** The recorded interviews with the children were transcribed. Both researchers were involved in the coding process of the transcripts from our interviewees. An interrater concordance was calculated, and it was found 95%. The categories that emerged were refined by the researchers and can be seen in Table 6 in the Results’ section. Content Analysis (CA) was used to analyze the interview transcripts. A Mann-Whitney U test was run with the categories (indicating different strategies) identified from the CA and gender.

**Results**

**Quantitative Findings**

**The Impact of Test Type, Gender and Age on Mathematical Test Scores**

The experimental tasks were used to examine if test-type (paper/iPad), gender and year group have an impact on mathematical test scores; a three-way ANOVA was utilized to examine this.

From Table 2, it can be seen that female students from all year groups had the highest iPad test scores compared to males, as seen by the means. Male students from all year groups had the highest paper test scores, as seen by the means, however, males’ paper test scores were more dispersed.

<table>
<thead>
<tr>
<th>Test Type</th>
<th>Gender</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>iPad</td>
<td>Female</td>
<td>10.59</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>9.75</td>
<td>1.21</td>
</tr>
<tr>
<td>Paper</td>
<td>Female</td>
<td>9.35</td>
<td>1.12</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>9.90</td>
<td>1.67</td>
</tr>
</tbody>
</table>
Table 3. Mean and standard deviations of test scores for year group and test type.

<table>
<thead>
<tr>
<th>Test Type</th>
<th>Year Group</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>iPad</td>
<td>Reception</td>
<td>9.82</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>Year 1</td>
<td>9.86</td>
<td>1.10</td>
</tr>
<tr>
<td></td>
<td>Year 2</td>
<td>10.75</td>
<td>1.14</td>
</tr>
<tr>
<td>Paper</td>
<td>Reception</td>
<td>9.45</td>
<td>1.04</td>
</tr>
<tr>
<td></td>
<td>Year 1</td>
<td>9.57</td>
<td>1.22</td>
</tr>
<tr>
<td></td>
<td>Year 2</td>
<td>9.92</td>
<td>1.24</td>
</tr>
</tbody>
</table>

resulting in a higher standard deviation. Assumptions have been met and the data is normally distributed.

From Table 3, it can be concluded that Year Group 2 had the highest iPad and paper test scores, as seen by the means. Year Group 2’s paper test scores were the most dispersed data set, as it has the highest standard deviation overall.

A three-way mixed ANOVA was run to understand the effects of test-type, gender and year group on test scores. Test scores were normally distributed, as assessed by Shapiro-Wilk’s test ($p > .5$), and there were no outliers in the data set, as assessed by inspections of boxplots. There was homogeneity of variances for both iPad test scores ($p = .343$) and paper test scores ($p = .751$), as assessed by Levene’s test for equality of variances. The three-way interaction between test-type, gender and year group was not statistically significant. There was a statistically significant two-way interaction between test-type and gender, $F(1, 31) = 9.043, p = .005$. All other two-way interactions were not statistically significant ($p > .05$). Statistical significance of a simple main effect was accepted at a Bonferroni-adjusted alpha level of .025. There was a statistically significant simple main effect of gender on the iPad test scores, $F(1, 31) = 7.259, p = .011$, partial $^2 = .190$, but not on paper test scores $F(1, 31) = 1.983, p = .169$, partial $^2 = .060$. All pairwise comparisons were performed for statistically significant simple main effects. Bonferroni corrections were made with comparisons within each simple main effect considered a family of comparisons. Adjusted $p$-values are reported. Mean iPad test score was higher for female students than male students, a mean difference of 0.913, 95% CI [0.222, 1.604], $p = .011$.

On-Line Parental Questionnaires: Exploration of iPad Usage at Home and the Relationship with Mathematical Test Scores

Parental questionnaires were utilized to investigate iPad usage at home such as: time spent on a tablet, applications used, if the child used an iPad etc. The on-line questionnaires were also used to examine if a relationship between daily time spent using an iPad at home and mathematical test scores (from experimental tasks) existed. A Spearman’s rank order correlation was used to determine this.

Three of the thirty-seven students (8%) did not have access to a tablet device in the home, as seen in Figure 1. Sixteen (43%) parents reported that their child uses a tablet device for 30–60 minutes per day, as seen in Figure 2. None of the parents declared their child spending over ninety minutes a day using technology. Three females (18%) used their tablet device for 60–90 minutes per day and only two males (10%) used their device for 60–90 minutes per day (see Figure 2).

Children’s main usage of iPad/tablet device at home, as reported by parents, is depicted on Figure 3, and it was for entertainment purposes, as 41% of children watched videos and 27% played games.

A Spearman’s rank-order correlation was run to assess the relationship between iPad mathematical test scores and daily time spent using an iPad at home in children aged 5 to 7 years old. Preliminary analysis showed the relationship to be monotonic, as assessed by visual inspection of a scatter plot (see Figure 4). There was a statistically significant, moderate negative correlation between daily time spent using an iPad at home and iPad mathematical test scores, $r_s (35) = -.453, p = .005$. Therefore, we can reject the null hypothesis and accept the alternative hypothesis. Therefore, we can confirm there is a significant relationship between iPad mathematical test scores and daily time spent using an iPad at home in children aged 5 to 7 years old.
Qualitative Findings and Exploration of the Different Strategies Employed by Children

Semi-Structured Interviews and Content Analysis

Interviews with the Children After the Completion of the Test to Explore the Strategies Employed to Solve the Mathematical Problems on Paper and the iPad

A Content Analysis (CA) (Neuendorf & Kumar, 2015) was performed on the semi-structured interviews. Each interview transcript was reviewed several times in order to create categories that explain the techniques children use to answer mathematical questions for: counting, shapes, addition and subtraction. Both researchers were involved in the coding process and an interrater concordance was ensured (95%). The categories that emerged were refined by the researchers and can be seen in Table 6. When these categories had been created, each answer given in the transcripts, fit into one of the three categories. The categories were ranked on how much skill is used, with one being the least skillful and three being the most skillful. Once the categories were established, each transcript was reviewed again, and the answers participants gave to how they answered mathematical questions were tallied next to the category. This data was then entered into SPSS, where the number of tallies for each gender and the percentage was presented in a table (see Table 4). A Mann-Whitney U test was run to investigate if there was a difference in methods used to answer questions between the genders.

What Strategies Did Female and Male Participants Used to Solve the Mathematical Problems on the Paper-Based Test and the iPad?

A Mann-Whitney U test was performed to determine if there was a difference in methods used by participants to answer mathematical questions between males and females on the paper-based test. Distributions of the methods for males and females were not similar, as assessed by visual inspections. Asymptotic significance was used to determine significance. Using mental representations to answer counting questions was statistically significantly different between males (mean rank = 23.52) and females (13.68), \( U = 260.5, z = 2.898, p = .004 \). Distributions of
the methods for males and females were similar, as assessed by visual inspections. No other statistically significant differences were found between female and male pupils in relation to the strategies employed when resolving the problems on paper. The results of Mann-Whitney U test presenting differences between female and male participants in terms of the strategies used to resolve the mathematical problems on the paper mathematical test are presented in Table 5 below:

A Mann-Whitney U test was performed to determine if there was a difference in methods used by participants to answer mathematical questions between males and females on the iPad-based test. Distributions of the methods for males and females were not similar, as assessed by visual inspections. Asymptotic significance was used to determine significance. The results of Mann-Whitney U test presenting differences between female and male participants in terms of the strategies used to resolve the mathematical problems on the iPad are presented in Table 6. Using mental representations to answer counting questions was statistically significantly different between males (mean rank = 22.57) and females (14.79) $U = 241.5$, $z = 2.462$, $p = .014$.

Distributions of the methods for males and females were similar, as assessed by visual inspections. We did not identify any other statistically significant differences between male and female pupils in terms of the strategies reported by the children when resolving the mathematical problems. Table 6 displays the mean rank and median scores for each strategy.
Figure 3. Main ways of using an iPad at home.

Figure 4. Daily time spent using an iPad at home being associated with iPad mathematical test scores.
Table 4. CA categories, showing how each gender answered mathematical questions on the paper test (N = 37).

<table>
<thead>
<tr>
<th>Mathematical ability</th>
<th>Rank</th>
<th>Category</th>
<th>Example answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Counting</td>
<td>1</td>
<td>Drawing on previous experience</td>
<td>I remember from math class/homework, I remember mom/dad telling me</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Using tangible objects</td>
<td>Counting on fingers, counting the pictures</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Mental Representations</td>
<td>Super-tising, counting in my head</td>
</tr>
<tr>
<td>Shapes</td>
<td>1</td>
<td>Drawing on previous experience</td>
<td>I remember mom/dad telling me, it is shaped like a football</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Using tangible objects</td>
<td>I looked at the picture</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Understand the qualities/characteristics</td>
<td>It has four sides, it is round, it has three edges</td>
</tr>
<tr>
<td>Addition</td>
<td>1</td>
<td>Drawing on previous experience</td>
<td>The teacher told us, I did it in class</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Using tangible objects</td>
<td>Counting on fingers, counting objects</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Mental representations</td>
<td>Super-tising, counting in my head</td>
</tr>
<tr>
<td>Subtraction</td>
<td>1</td>
<td>Drawing on previous experience</td>
<td>The teacher told us, I did it in class</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Using tangible objects</td>
<td>Counting on fingers, counting objects</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Mental representations</td>
<td>Super-tising, counting in my head</td>
</tr>
</tbody>
</table>

Table 5. Mean rank/median scores of each CA category for both genders from the Mann-Whitney U test for the paper mathematical test.

| Sub-Category | CA Category | Mdn Score | Mean Rank | |
|--------------|-------------|-----------|-----------|
|              | Female      | Male      | Female    | Male     |
| Counting     | Drawing on previous experience | 1 | 1.45 | 22.32 | 16.18 |
|              | Tangible objects | | | 13.68 | 23.52 |
|              | Mental Representation | | | |
| Shapes       | Drawing on previous experience | 1.41 | 1.25 | 21.91 | 16.52 |
|              | Tangible objects | | | 15.56 | 21.93 |
|              | Understand the qualities/characteristics | | | |
| Addition     | Drawing on previous experience | 1.24 | 1.45 | | |
|              | Tangible Objects | 1.76 | 1.15 | | |
|              | Mental Representation | 0.94 | 1.55 | | |
| Subtraction  | Drawing on previous experience | 1.18 | 1 | | |
|              | Tangible Objects | 1.59 | 1.15 | | |
|              | Mental Representations | 1.06 | 1.55 | | |

Table 6. Mean rank/median scores of each CA category for both genders from the Mann-Whitney U test for the iPad mathematical test.

| Sub-Category | CA Category | Mdn Score | Mean Rank | |
|--------------|-------------|-----------|-----------|
|              | Female      | Male      | Female    | Male     |
| Counting     | Drawing on previous experience | | | 20.35 | 17.85 |
|              | Tangible objects | | | 21.50 | 16.88 |
|              | Mental Representation | | | 14.79 | 22.57 |
| Shapes       | Drawing on previous experience | | | 18.38 | 19.52 |
|              | Tangible objects | | | 17.50 | 20.27 |
|              | Understand the qualities/characteristics | | | 20.74 | 17.52 |
| Addition     | Drawing on previous experience | | | 15.21 | 22.23 |
|              | Tangible Objects | | | 23.97 | 14.78 |
|              | Mental Representation | 1.24 | 1.15 | | |
| Subtraction  | Drawing on previous experience | | | 16.26 | 21.32 |
|              | Tangible Objects | 1.65 | 1.75 | | |
|              | Mental Representations | | | 21.53 | 16.85 |
Discussion and Conclusions

Several key findings have emerged from this study. Firstly, test-type and gender have appeared to impact mathematical test scores for primary school children. More specifically, female participants scored higher on average on the iPad mathematical test. Even though year group was not found to impact test scores, this study discovered that the year groups sub-scores varied significantly between iPad and paper tests. Year 2 had the most dispersed data sets compared to the other year groups. In addition to this, Year 2 performed highly in the sub-categories of addition and subtraction for both iPad and paper tests and performed poorly in counting and shapes. Reception and Year 1 performed similarly in all the sub-categories for both tests, except for Year 1 performing highly in shapes for the iPad test. From further analysis it was found that there is no significant difference in iPad and paper test scores between males and females.

Analysis of the parental questionnaire illustrated that the majority of participants: owned a tablet device at home, spent 30–60 minutes per day using their tablet and predominantly used it for watching videos/playing games. Educational apps and homework tasks were the least likely to be used by participants. A significant correlation was found between iPad mathematical test scores and daily time spent using an iPad at home. The correlation showed that as daily time spent using an iPad increased, then iPad mathematical test scores decreased. This relationship occurred especially after a participant used their iPad for more than 60 minutes a day.

Content Analysis revealed that using tangible objects was favored when answering addition and subtraction questions. Whereas mental representations were favored for counting questions; drawing on previous experience was favored for shape questions. Further analysis found there were only significant differences in mental representations for counting between males and females and differences in tangible objects for addition between males and females. These indicated that males used mental representations more often for counting and females used tangible objects more often for addition. A discussion of the research questions is following:

Discussion of Research Questions

The research question: does test type, age, gender influence mathematical test scores, has been explored in this study. As discussed, it was found that test-type and gender did influence mathematical test scores, whereas age did not. The results of the analyses showed that there was a distinct relationship between females and iPad mathematical test scores, as they performed best on this test. This is an interesting finding, as females have been seen to underperform males in mathematics over the years of research (Contini et al., 2017). An answer for this may be due to the fact that other research has suggested that females have been known to adopt a visual style to learning, receiving higher benefits from this (Pruet et al., 2016). Females have reported using educational apps more frequently than males (Nang et al., 2015), therefore, these previous results may provide an explanation of the current study’s finding that females have higher scores on technology based mathematical tests. Furthermore, test type and the testing procedure being influential on mathematical test scores can be validated by the fact that using technology in the classroom promotes interest in the specific subject it is being used in (Uzoğlu & Bozdoğan, 2012), resulting in higher engagement and academic attainment (Kalati et al., 2022) and making the use of tablets in math testing another option and application of technology in learning and assessment.

The hypothesis if a child uses an iPad/tablet device frequently, then their test scores will be higher, was not supported by the findings of the study; it was found that iPad test scores decreased, when daily iPad usage increased, especially when usage was over 60 minutes per day. These results shine light on why the World Health Organisation suggests that children should only have 1 hour of ST in 24 hours (WHO, 2019). It is possible that children’s increased use of screen time may hinder their cognitive development or concentration on the task given to them and their attention span (Herodotou, 2018).
Previous literature explained that children’s ST is constantly increasing and the growing dependency on tablet devices is becoming a societal need and issue that needs further investigation (Hadlington et al., 2019; Romero-Tena et al., 2022).

Female participants were reported to use their tablet device for longer time than males. This is relatively novel, as previous research has found that technology is vastly used more by males (Colley, 2003; Ferguson, 2017; McKenney & Voogt, 2010). Palaiologou (2016) and Taufik et al. (2019) propose that, in European countries, children, both boys and girls, use digital technologies such as computers, tablets, smartphones, apps and gadgets more often, generating effects on their behaviors. The digital environment of children is increasingly becoming rich and diverse, and studies about this topic have documented its evolution by countries and technology (Konca & Koksalan, 2017). Konca (2021) pointed that the prevalence of digital technologies is a characteristic of the daily life of families now, supporting learning both at home and at school. Overall, the results do not support the hypothesis proposed, regardless they do help justify guidelines for technology usage for children. Kalati et al. (2022) have proposed that further investigation is needed to facilitate the identification of the factors and conditions accounting for the optimal learning outcomes during young children’s learning with touchscreens and the optimal time of exposure to touchscreens.

Axell and Boström (2021) had pointed out that technology is strongly connected to the female/male dichotomy, reporting gender differences in the attitudes toward technology. The “norm” (hidden) is to consider that men can have greater knowledge in the technological scope, posing a barrier for women to approach this field (Axell & Boström, 2021). This position was not confirmed from the findings of the study as girls did use iPads at home and had higher performance in the mathematical problems on iPad. In this whole process of development and acquisition of norms and habits, families also seem to generate positive beliefs about the use of technologies by their children, and they believe that learning with technologies provides new and enjoyable learning opportunities for children (Nikolopoulos, 2020).

The hypothesis that female participants will rely on tangible objects to answer mathematical questions and males may rely on mental representations, was partially supported by the current study’s findings. It was found that females used tangible objects for addition questions and males used mental representations for counting questions. This illustrates that each gender utilizes different methods for certain mathematical abilities. Females may use tangible objects more for addition, as they have a visual learning style (Pruet et al., 2016), therefore, they would need physical objects to interpret and understand what is being asked of them. The results can also be explained by research based on gender technology usages. Females use apps such as coloring and style creation, where physical objects are being used and they may learn to rely on the physical objects when cognitive abilities are in use. Males use strategic apps, where mental thinking and processing occurs, therefore they rely on using mental representations when using cognitive abilities (Marsh et al., 2018).

There were several results which were found but not hypothesized. Year 2 had the most dispersed data for iPad and paper test scores. This may be because two participants who did not own a tablet device at home were pupils in Year 2. They may not have been familiar with using an iPad, therefore results could have been impacted by this. Child age and previous experience/familiarity with touchscreens can affect children’s learning with the use of touchscreens (Kalati et al., 2022; Schmid et al., 2023).

**Limitations and Future Research**

The limitations of the present study include sample size, environmental conditions and time constraints. Sample size could have been larger, and the analyses would have benefitted from a wider sample. Some data was not normally distributed and non-significant, this may have occurred due to a lower number of participants. The target population was children between the ages of 4–7 years old. Due to the age of the participants, legally parents need to provide consent for the children to participate. Many parents may not have had the time to read
through the participation sheet and consent form, in order to give consent. Another limitation of the current study is that we do not know to what degree the children had access to technology in their schooling. Did some teachers use technology in their curricula? Did other children not have this access? It seems that prior experiences with technology in schools would play a major role in how they responded to technology-based testing procedures in our study. Further research is required to better understand the impact of testing procedures in math and factors influencing how young children answer the questions posed and strategies employed when answering problems on counting, knowledge of shapes, addition and subtraction.

It would be inevitably interesting to administer the research tools in a possibly different setting or conditions, when it was not perceived as a test but as a learning activity or play-based learning. Future research on modes of assessment in actual classrooms could also result in valuable information on this topic.

**Implications for Practice And/Or Policy**

Digital testing may enable a better investigation of mathematical skills in the first years of schooling and of differences between males and females’ responses to solving mathematical questions, which then could be used to tailor the curriculum. The guidance for technology usage can be confirmed through the investigation of home usage and school ability. Our findings have potentially implications for different stakeholders by giving them insights into the impact of testing procedures and practices on pupils’ performance in math with the use of touchscreen devices and on young children’s learning and attainment under different testing conditions. In future classrooms, it can be explored how learning and assessment of mathematical skills could be implemented with the use of iPads with young children in different learning environments.

**Statements on Open Data, Ethics and Conflict of Interest**

There was no conflict of interest in this study and this work has not been previously published. The research has been conducted following the ethical guidelines and principles of Birmingham City University and the British Psychological Society and has been approved by the Departmental Research Ethics Committee and the Chair Dr Emma Bridger (DREC Code: PSY_BSc_Jun19_004). The personal data of the participants is not available for this study, as all data was anonymized through pseudonyms and then the personal data was deleted to protect the participants due to their age. Also, participants were made aware that their participation was voluntary and that they could withdraw without parents’ consent. The study excluded students with learning difficulties, as completing the tasks would have caused unwanted stress. The data of the study can be shared by the authors upon request; the data that support the findings of this study are not openly available as the research study recruited young children and participating parents and children of this study did not agree for their data to be shared publicly, so supporting data can be made available upon request.

**Disclosure Statement**

No potential conflict of interest was reported by the author(s).

**ORCID**

Olga Fotakopoulou [ID](http://orcid.org/0000-0001-5041-7021)
References


