Article:

Sunde, PB, Sunde, P and Sayers, J orcid.org/0000-0002-9652-0187 (2020) Sex differences in mental strategies for single-digit addition in the first years of school. Educational Psychology, 40 (1). pp. 82-102. ISSN 0144-3410

https://doi.org/10.1080/01443410.2019.1622652

Sex Differences in Mental Strategies for Single-Digit Addition in the First Years of School

*Pernille B. Sunde[a,b], Peter Sunde[c], and Judy Sayers[d]

[a]Danish School of Education, Aarhus University, Copenhagen, Denmark; [b]VIA University College, Aarhus, Denmark; [c]Department of Bioscience, Aarhus University, Aarhus, Denmark; [d]School of Education, University of Leeds, Leeds, United Kingdom.

* Pernille B. Sunde, Department of Educational Psychology, Aarhus University, Copenhagen, Denmark. e-mail: pebs@edu.au.dk

ORCiD:
Pernille B. Sunde: http://orcid.org/0000-0001-5009-0233
Peter Sunde: https://orcid.org/0000-0002-7485-037X
Judy Sayers: http://orcid.org/0000-0002-9652-0187

Word count: 8198

Acknowledgements

We would like to thank The National PhD Council in Educational Research, Denmark for supporting the collection of this data.

Sex Differences in Mental Strategies for Single-Digit Addition in the First Years of School

Strategy use in single-digit addition is an indicator of young children's numeracy comprehension. We investigated Danish primary students' use of strategies in single-digit addition with interview-based assessment of how they solved 36 specific single-digit addition problems, categorized as either 'error', 'counting', 'direct retrieval' or 'derived facts'. The proportional use of each strategy was analysed as multi-level functions of school age and sex. In a first study (260 interviews, 147 students) we found decreasing use of counting and increasing use of direct retrieval and derived facts through year 1-4, girls using counting substantially more and the other two strategies substantially less than boys, equal to more than two years' development. Similar results appeared in a subsequent study (155 interviews, 83 students), suggesting that the pattern is pervasive in Danish primary schools. Finally, we ask whether sex differences in strategy use is generally under-reported since many studies do not explicitly address them.

Keywords: sex differences; single-digit addition; strategies; years 1 to 4; mathematics.

Introduction

Appreciating sex differences of student competences in mathematics has long been a discussion across time and across borders (Carr, Steiner, Kyser, & Biddlecomb, 2008; Fennema, Carpenter, Jacobs, Franke, & Levi, 1998; Shen, Vasilyeva, & Laski, 2016). The topic continues to attract attention from education in general, mathematics education in particular, as well as more traditional psychological perspectives (Benbow, 1988; Browne, 2002; Connellan, Baron-Cohen, Wheelwright, Batki, & Ahluwalia, 2000; Halpern et al., 2007; Maccoby & Jacklin, 1974; Pinker, 2002). Research findings also attract attention for their diversity and contradiction, where either boys are found to outperform girls (Fennema et al., 1998; Pinxten, Marsh, De Fraine, Van Den Noortgate, & Van Damme, 2014), or where girls show higher achievement than boys (Voyer & Voyer, 2014).

Research in sex differences have been seen between age and achievement levels within different achievement groups (e.g. male overrepresentation in high achieving and low achieving groups (Penner & Paret, 2008). Moreover, high parental education has been reported to advantage young male students entering kindergarten, in not only middle and upper class families, but noted also in Asian families. All of which exhibit the largest male advantage at the top of the distribution. In contrast and of note, Penner and Paret (2008), report that Latino kindergarten girls have an advantage over boys at the top of the distribution. However, the effect sizes of such studies are often challenged (Hyde, 2005; Hyde, Fennema, & Lamon, 1990; Penner & Paret, 2008; Shen et al., 2016; Spelke, 2005). For example, Else-Quest, Hyde, and Linn’s (2010) meta-analysis of TIMSS and PISA data showed very low effect sizes, but, they also suggested that there was a higher variability between the countries in those effect sizes in specific domains of mathematics. An example of the influence of educational and cultural context on young students' arithmetical strategies, is the study of Shen at al. (2016). They found that sex differences in the use of strategies for mixed- and double-digit addition and subtraction was present for American and Russian first graders but not for Taiwanese, indicating the importance of educational and cultural context.

Sex differences in mathematics ability related to number and arithmetic may in theory also be linked to underlying differences in cognitive factors that are known to correlate with mathematics ability, e.g. mental rotation skill (e.g. Laski et al., 2013; Moè, 2018) and working memory (e.g. Geary, Hoard, & Nugent, 2012). Sex differences in for example mental rotation skill has been shown for children as young as 4½ years (Levine, Huttenlocher, Taylor, & Langrock, 1999) as well as college students (Casey, Nuttall, Pezaris, & Benbow, 1995). Sex differences in cognitive factors does on the other hand not necessarily link to sex differences in calculation performance and number competences. For example, Moè (2018) found no sex differences in numbers and arithmetic ability of eight and ten year old children despite that boys scored higher than girls in mental rotation (2-D and 3-D), a factor that in turn correlated with mathematics performance at ten years' of age.

In recent years, attention has focussed on the importance of identifying individual differences children display in the early learning of number, in order to inform appropriate intervention planning and teaching for children (Dowker, 2005). Research also points to identifying what aspect of mathematics children struggle with. For example, Hornburg, Rieber, and McNeil (2017) analysed data from 14 studies (including 960 second and third graders) on boys' and girls' understanding of mathematical equivalence. They reported that in arithmetic, girls are more likely to construct narrow knowledge, i.e. memorise taught procedures and rely on that knowledge when encountering novel problems. Consequently, this causes differences between them and boys who demonstrate broader knowledge of number.

In light of the above, we appreciate that there are no simple distinctions between sex differences, or when students might present differences in different mathematical components. Thus, explanations for sex differences can be found in individual differences as well as sociocultural factors. Indeed, there is already a significant amount of research in these areas, which is receiving increasing recognition (e.g. Hirnstein, Andrews & Hausmann, 2014). The study presented here sought to investigate how young boys and girls in years 1 to 4 develop their strategies in single-digit addition in a Danish context.

Denmark

The study is of interest to a wider audience for a number of reasons. Firstly, little is published on Danish elementary school mathematics, its content and pedagogy. Secondly, according to Selter (2001, p. 147), standard methods of calculation are given a marginal position in the arithmetic curriculum focus. Lastly, and not insignificant, the Danish curriculum, and its assessment framework, are at the root of the international PISA testing instrument.

The teaching of mathematics at compulsory school in Denmark (age 6-15), has undergone much self-scrutiny in recent years, and teacher education has undergone some major transformations since 2001. A decade later, Niss and Højgaard (2011) recommended that the Ministry of Education “ensure that *teacher training* at all educational levels under the jurisdiction of the Ministry be designed and structured so that future teachers are equipped with the mathematical, didactic and pedagogical competencies presented in this report.” (Section 1.5, p. 194). These competencies are now part of the official Danish curriculum, but Danish teachers enjoy great autonomy in the classroom (Skott, 2004) and are often reported to uncritically trust textbook content to deliver curriculum outcomes (Bremholm & Skott, 2019). The study here seeks to illuminate what arithmetic strategies Danish children are working with in single-digit addition. The operation that is reported to be the foundations of primary mathematics instruction (Carpenter & Moser, 1984).

Strategies in Single-Digit Mental Addition

Carpenter and Moser (1984) identified three basic levels of addition strategies based on 1) direct modelling strategy with fingers or physical objects where counting sequences are used by children to count all concrete objects; 2) Counting strategies, where children count on from the first or the larger number, used in developing mental counting and from counting all starting with the first addend to counting all starting with the larger addend and finally to counting on from the larger addend, in other words, counting-all is replaced by counting-on strategies. These strategies have, and since been reported to be used for some time before using strategies based on recalling number facts (Sarama & Clements, 2009). Finally, 3) Number fact strategies which involve either a) direct retrieval, or b) derived fact strategies. Direct retrieval (fact retrieval or just retrieval) is the direct recalling of the solution to an addition question from memory (Carpenter & Moser, 1984). These known facts are an important prerequisite of derived fact strategies, the focus of many recent studies (e.g. Dowker, 2014; Laski, Ermakova, & Vasilyeva, 2014). Derived fact strategies, where number facts are derived from a recalled number fact (Carpenter & Moser, 1984, p. 181), are also known as decomposition or partitioning strategies. These strategies build on known facts and involve transforming the addition question into more simple questions. Derived fact strategies fall into two distinct groups based on known facts or, base-10 knowledge

(Dowker, 2014). An example of the former is to solve the unknown question of 7+8 by transforming into the known sum of 7+7 and then add 1. An example of the use of base-10 knowledge is solving the same question 7+8 by partitioning 7 into 2 and 5, adding 8 and 2 to get 10, and then add the remaining 5 to get the final sum 15.

Derived fact strategies has been associated with conceptual understanding (Canobi, Reeve, & Pattison, 1998) and number knowledge, e.g. base-10 knowledge (Laski et al., 2014) and are thus considered to be more advanced strategies than counting.

Strategy Use in Mental Addition and Relationship with Mathematics Achievement

The use of mental strategies in arithmetic has been shown to influence development of mathematics competence (e.g. Carr et al., 2008; Dowker, 2014; Fennema et al., 1998; Ostad, 1997) and different strategy use is thus a valid predictor of both later mathematical achievements and difficulties (Carr & Alexeev, 2011; Gersten, Jordan, & Flojo, 2005; Ostad, 1997; Price, Mazzocco, & Ansari, 2013). Yet, in the Nordic context Ostad's (1997) longitudinal study has shown that children without difficulties use "a richness of strategies" (p. 355) comprising of both direct retrieval and derived fact strategies, which increases as the numbers used confidently increases as they progress through school. Thus, children using retrieval strategies at an early age (6-7 years) are less likely to develop difficulties in mathematics. However, children with difficulties were found to use fewer and inadequate strategies, primarily backup strategies (counting), and fix their choice of strategies early in school with little subsequent progression (ibid.).

The debate of individual differences is related to strategy use and achievement increases, as more recently Bailey, Littlefield, and Geary (2012) reported that the preference for and skill at using a specific strategy related to a feedback loop, (a process for checking for and affirming understanding, thus can be negative as well as positive outcomes), predicts the later skill. In this case, early preference of strategies used for addition successfully which in turn predicts later preference, and so forth. Students preferring counting strategies will get less practice in using derived fact strategies and thus working flexibly with numbers. Consequently, they will have fewer opportunities to develop the more complex understanding necessary for advanced arithmetic (Gersten et al., 2005). However, Bailey et al. (2012) also found a variation in such influences between girls and boys across grades 1-6 in North America, indicating the importance of identifying such preferences in other contexts and cultural settings.

Sex Differences in Use and Development of Strategies in Addition

When considering strategies in addition specifically, research evidence continues to be ambiguous. Sex differences in use and development of strategies in single-digit addition has been addressed in several studies from different western countries.

In a study on 6 and 7 year olds Dowker (2009) found no sex differences in addition performance or use of derived fact strategies. However, other studies report on consisting sex differences (e.g. Bailey et al., 2012; Carr & Alexeev, 2011; Carr & Davis, 2001; Imbo & Vandierendonck, 2007; Paul & Reeve, 2016; Shen et al., 2016). Bailey et al. (2012) found that boys showed a bias towards the use of retrieval and girls showed a bias for the use of counting procedures from years 1–6, a result that was consistent with previous findings by Carr and Davis (2001), Carr and Jessup (1997), and Geary, Bow-Thomas, Liu, and Siegler, (1996). The developmental pattern for skilled retrieval is more complex. For example, in first grade, boys are reported to have a higher preference for retrieval than girls, but the difference in accuracy was insignificant at this age. But by the second grade, although girls were still shown to retrieve less than boys they were more accurate than boys when they did use it (Bailey et al., 2012). This reflects Siegler's (1988) research findings, resonating the identification of Carr and Alexeev's (2011) 'perfectionists': girls were less likely to adopt a retrieval strategy than boys, as they want to be sure of accuracy, even though this strategy is more time-consuming. Contrastingly, this early preference by boys, according to Bailey et al. (2012), was found to increase steadily from first to sixth grades, and at this point out-performed girls. It is important to note that Bailey et al.'s (2012) results indicate, that ability influences early skilled retrieval, but both practice and skill influence each other in a feedback loop later in development. Thus, there appears to be a change within grades 1-6.

Boys' early advantage for accurate retrieval was not due to IQ or 'central executive' according to Geary et al. (2012), and research suggests that boys generally take more risks than girls, especially in situations where their performance will be socially evaluated (Geary, 2010). Boys like to express themselves more freely in open settings that might encourage a competitive environment. Peterson and Fennema (1985) reported that boys tended to enjoy getting the answer before anyone else. However, Bailey et al. (2012) reported that boys had shorter reaction times than girls in grade one, but that girls begin to catch up with them, whereas boys' reaction times did not increase in the same way. They suggest this is because the competitiveness is about being seen to be first to answer, rather than accuracy. Which confirms previous studies that suggested that this competitive perspective may harm boys' addition accuracy in the early grades, but this early preference may put them at an advantage later on (Royer, Tronsky, Chan, Jackson, & Marchant, 1999).

The Current Study

This study parameterizes the use and development of strategies in single-digit addition by gender over time (age). We focus on the first years of school in a four year longitudinal perspective. Furthermore, we investigate the different strategy categories (counting, direct retrieval and derived facts strategies) separately.

Our research aim is to quantify the variation and development in students' use of different strategies for single-digit addition during the first four years (years one to four) of Danish school, including any possible sex differences. We apply a multilevel model approach to ensure the generalisability of the findings and test the model on an independent dataset.

Methods

Participants

Data consist of information from assessment interviews on 230 Danish primary students' unprompted strategy use for single-digit addition from two independent studies: A and B, where study B was used to validate the results from grade 1, in study A. Informed written parental consent was obtained for all students. Students were informed orally about the study and that participation was voluntary.

Study A

Study A consisted of 260 assessment interviews conducted on 147 students (77 girls) from eight classes from a single school from year one to four (age 7 to 11). Three classes were tested thrice in subsequent years; two classes twice; and three classes once (Table 1).

INSERT TABLE 1 ABOUT HERE

Study B

Study B consisted of 155 assessment interviews from 83 students (46 girls; age 7) from six classes from three other schools. Students were interviewed in the first and second half of year one (Table 2).

INSERT TABLE 2 ABOUT HERE

Assessment

Interviews on unprompted strategy use. Strategy use was monitored in one-to-one assessment interviews by presenting the student with flashcards with the 36 addition tasks with numbers 2-9, including doubles. The addition tasks were presented with sums less than 10 first and then with increasing sum and difficulty. This was to avoid less confident students, as reported in Dowker 's (2003) study, giving up or only using counting because they were presented with difficult questions (e.g. 8+7) at the beginning of the interview. The students were not provided with any manipulatives or paper and pencil. The interviews lasted 10-30 minutes and were performed in a quiet room at the school. The solving time for the individual addition task was not measured.

In light of the children being so young, the interview script was cross checked for clarity and appropriateness with researchers and teachers of the children. The researcher stated to each student: "I will show you some single-digit addition tasks. First, I would like you to find the answer to the task, and then we talk about how you found the answer. There are many ways to find the answer to an addition task. Sometimes you might know the answer or count or perhaps you use other tasks to find the answer. I am interested in knowing how *you* find the answer." Then the student was presented for a flashcard, e.g. 4+5, and the interviewer asked: "what is the answer to four plus five?”. If the student did not give an explanation following the answer, the interviewer asked: “how did you find the answer?” and if further prompting was needed: “did you count or did you just know the answer, or did you use some other tasks you know to find the answer?”.

Categorisation of strategy use. If the student gave up or miscalculated an answer, this was categorised as Error. Strategy use for incorrect answers were not categorised. Correct answers were categorised based on the student’s self-report of strategy use and observations by the interviewer (e.g. visible signs of finger-counting or lip movements). The validity of young students self-reported strategy use has been established in other studies (e.g. Canobi et al., 1998) based on comparison with solution time. Strategy use of correct answers were categorised in Counting all (counting both addends and then all together), Counting on (counting on from one of the addends), Direct retrieval (reported just knowing the answer), Derived fact using addition (decomposing addends and calculating answers using automatized sums with subsequent use of addition e.g. $4 + 5 = 4 + 4 + 1$), and Derived fact using subtraction (decomposing addends and calculating answers using automatized sums with subsequent use of addition e.g. $4 + 5 = 5 + 5 - 1$). We did not distinguish between counting on fingers, verbal counting or self-report of mental counting. Initial analyses showed that the categories counting all and both derived fact categories were used relatively infrequently by the students. It was therefore difficult to establish robust statistical models for these rare categories. Therefore, we pooled the counting and derived fact strategies respectively to 1) provide more reliable results and 2) to reduce complexity of the analysis. The pooled categories of strategy use for correct answers used in the analysis was thus Counting, Direct retrieval and Derived fact.

Data Analysis

Development of strategy use (Study A). Variation in the frequency by which a given solving strategy was used by students in study A (binary response variable: e.g. whether direct retrieval was used as opposed to all other strategies) to solve a specific addition task (e.g. ‘5+6’) as a function of school year (number of years the student has attended school), sex and the interaction was analysed by means of Generalised Linear Mixed Models (GLMM: GLIMMIX procedure in SAS 9.4, SAS Inst. (Littell, Stroup, Milliken, Wolfinger, & Schabenberger, 2006)) with a logit link function and binomial error distribution.

We used the answers of the individual addition task (36 per assessment interview) as observation unit. The entire data set thus consisted of a total 9360 addition tasks based on 260 assessment interviews, nested within 147 students from eight classes. To account for the expected dependency of observations within this hierarchical data structure, we stated the following random effects in the model: (1) assessment interview nested within student ID (accounting for dependency of strategy use among task cases within the interview), (2) student ID nested within class (accounting for random variation between students within a class), and (3) class ID (accounting for random variation between classes). Finally, the identities of the specific addition task (e.g. ‘2+2’, ‘2+3’ etc.: 36 different task IDs in total) were also stated as random effects to account for any variation in problem solving strategy related to the specific task.

As fixed effects, we included sex (categorical variable) and school age (covariate representing the number of years the student had received school education: for example, if school year runs from 8 August to 26 June, a student tested in year one on 23 October had calculated school age of 1.24 and when tested again on 29 March in the same year a school age of 1.72) and the interaction term between these two factors.

The statistical significance of the fixed effects was based on p-values derived from t- and F-statistics of the models (the two test statistics rendered the same results, except a single case declared as such where the model had become over-parameterised because of inclusion of a superfluous interaction term, resulting in an unreliable t-statistics because of a poorly estimated SE). We used $p < 0.05$ as significance level for whether we would draw inferences from a result. Degrees of freedoms in the GLMM were calculated with the Satterthwaite approximation (Littell et al. 2006). Model predictions (predicted universal probability of using a given strategy for a given sex at a given school age with 95% confidence errors) were obtained as least square means.

Effect sizes were expressed as the odds ratios of the difference in strategy use by mid year one (school age = 1.5), derived as exp(b), where b is the coefficient of the gender effect on logit scale by school age = 1.5 (extracted from SAS as a least square means estimate specific to school age = 1.5).

Validation of results (Study B). The same model structure was used as in study A. Model predictions (least square means estimates) for the two studies were compared for mid-year one (least square means predictions for year =1.5). The difference between estimates obtained in study A and B were tested on the t-statistics of the difference in parameter values ($t_{(dfA+dfB)} = [b_A - b_B]/[SE_{b(A)}^2 + SE_{b(B)}^2]^{0.5}$).

We also split the analysis in study B on the six different classes in order to visualize predictions if estimated separately for each class. We furthermore tested explicitly for between-class variation in sex specific strategy use by entering class ID as a fixed effect. Likewise, we tested for variation between classes in sex specific strategy use by entering a sex-by-class interaction term in the model's fixed effect statement.

Post-hoc analysis: counting divided on counting all and counting on. After the overall result from study A had been replicated in study B, on the combined (and statistically more powerful) dataset, we quantified the sex differences in counting all and counting on, i.e. the two sub-categories comprising the overall category counting (same model structure as above).

INSERT FIGURE 1 ABOUT HERE

Results

Study A

The frequency by which all four strategy categories were used varied as function of school year, as 'error' and 'counting' were used less and the more advanced strategies 'derived fact' and 'direct retrieval' were used more with increasing school year (Figure 1, Table 3). Girls used "counting" significantly more (F:M odds ratio [95%CI] by mid year one = 2.9 [1.9-4.3]) and direct retrieval (0.57 or 1:1.8 [0.45-0.73]) and derived fact strategies (0.30 or 1: 3.3 [0.16-0.58]) significantly less often than boys. For counting, the effect size of sex corresponded to 2½ years development (Table 3); or in other words: the average use of counting by boys in the start of year one equalled the use of girls by mid year three (Figure 1). Same patterns applied for direct retrieval and derived fact (Table 3, Figure 1). On arithmetic scale, all sex differences diminished with school age. Hence, by year four, girls' strategy use in single-digit addition nearly equalled that of the boys.

The category error showed no sex differences (Table 3). For counting and derived fact strategies, sex differences were statistically significant as main effects only

(Table 3). This means that the relative difference between boys and girls was constant irrespective of school year on logit scale, but diminished with increasing school age when back-transformed to arithmetic scale (Figure 1) due to a reduced space within which the relative difference between sexes could be expressed in terms of percentage points when the proportional use of these strategies converged towards 0 and 1, respectively (boundary effect).

For direct retrieval, a statistically significant sex-by-school year interaction (Table 3) suggested that not only absolute (arithmetic scale: Figure 1) but also the relative sex difference (logit scale) in the inclination by which boys and girls used this strategy decreased with increasing school year.

In addition to the afore described mean trends in strategy use as function of school year and sex, considerable variation in strategy use could be attributed to the individual test situation (nested within student) and student identity (nested within class). This variation is apparent as the scatter around the predicted functions of the mean (Figure 1) as well as covariance parameter estimates of test and student identity as random effects (Table 3). In contrast, modest covariance parameter values of class identity relative to those for test and student identities indicated that most individual variation in strategy use occurred on test and student level rather than at class level. Lack of statistical significance of the class identity covariance parameters (Table 3) further suggested that the influence of class environment on strategy use was minor if present at all. The magnitude and statistical significance of the covariance parameter values of addition task identity also suggested that considerable variation existed in the probability by which the strategy in question was used to solve the individual addition tasks.

INSERT TABLE 3 ABOUT HERE

INSERT FIGURE 2 ABOUT HERE

Study B

Study B produced similar predictions for the difference between boys and girls in strategy use as well as strategy use specific to sex (Figure 2, Table 4 and 5) to study A: girls used counting significantly more often and derived fact and direct retrieval significantly less often than boys (Figure 2, Table 4). With exception of 'error' that occurred more often in study A than study B (Figure 2, Table 4), the proportional use of the different strategies was also equal in the two studies (Figure 2, Table 4).

The sex specific differences in the proportional use of 'counting', 'derived fact' and 'direct retrieval' estimated were also apparent if estimated separately for the six different classes in study B (Figure 2), and the magnitude of the sex effects did not vary significantly between the classes (Table 5).

Post Hoc Analysis of Counting All and Counting On

On the combined data from study A and B, by mid-year one, the sex difference (F:M odds ratio) in counting all and counting on was 3.2 (95%CI: 1.5-6.8, $t_{213.1}$=3.05, P=0.0027) and 2.1 (95%CI: 1.5-3.2, $t_{188.6}$=3.27, P=0.0014), respectively.

INSERT TABLE 4 ABOUT HERE

INSERT TABLE 5 ABOUT HERE

Discussion

This study demonstrates substantial mean differences in development patterns in the use and development of strategies for single-digit addition of girls and boys in the first years of Danish primary school, with girls using counting strategies 3 times more and derived fact strategies 3 times less often than boys by mid year one. This difference that was equal to girls being 2-3 years behind boys in development of strategy use was highly statistically significant and consistent across schools and classes, suggesting that it was general in the Danish primary school context. The girls' predisposition to use counting strategies more often than did boys applied for the least advanced counting all strategy as well as for the slightly more advanced counting on strategy.

Our results appear to be comparable to those of Bailey et al. (2012), where sex differences were found in the use of direct retrieval equal to approximately two school years in a longitudinal study (year 1-6) in USA. Our results also resembles that of Carr and Jessup (1997) who found that boys in year one in USA used retrieval more often than did girls. However, unlike us they found an increasing sex difference from October (where boys and girls used retrieval equally) to April (where boys were using retrieval almost twice as often as the girls). In Flemish Belgium, Imbo and Vandierendonck (2007), found that sex differences explained 4-5% of the variation in year four to six students' frequency of use of retrieval, also after controlling for age, arithmetic skill, working memory load and processing speed. However, the estimated difference in sex-specific strategy use (e.g. odds ratios) was not reported. Finally, Geary et al. (2012) in USA (kindergarteners) and Paul and Reeve (2016) (Australia, kindergarten to year 3) reported that boys differed significantly from girls in arithmetic strategies in the same direction as we observed, but gave no details on the size of these differences.

Other studies have reported differences in strategy use to be related to cognitive factors (Foley, Vasilyeva, & Laski, 2016; Laski et al., 2013), and indeed cultural factors such as different instructional practices, demonstrated to be apparent in several cross-country studies (e.g. Hyde & Mertz, 2009; Penner & Paret, 2008; Shen et al., 2016). Relating individual variation (including sex differences) in strategy use to other parameters such as general mathematical achievement or cognitive development is beyond the scope of this paper (a thorough analysis of data from study B that also include longitudinal data on mathematical achievements by the end of fourth grade will be published later). Nevertheless, we can reveal that year one students' general mathematical achievement was similar for boys and girls, but correlated negatively with proportional use of error and counting all but positively with proportional use of derived facts, suggesting some association between strategy use and general mathematical achievement in year one.

Implications

The marked sex-difference in development of strategy may have implications for research as well as teaching practice. Regarding research of variation and development of strategy use, our results suggest that sex may be a significant component of variation that should not only be tested for (or neutralised by balanced study designs), but incorporated in the statistical analyses instead of appearing as apparent statistical background noise. Sex differences in strategy use (or any other behaviour or skill) may also be of interest in its own right in order to identify differences between boys and girls in a given age or cultural context.

For teaching practice, knowledge of the existence of substantial variation in different students' development of strategy use, much of which connecting to the individual's sex may also be of importance. For instance, students who overly rely on counting in single-digit problem solving (of which girls are overrepresented) may need explicit teaching and encouragement to memorise number facts and patterns in number bonds. Thus, if instruction in the early years of school focus on counting strategies and procedural knowledge in the teaching of arithmetic this would disadvantage the girls as they are more prone to construct narrow knowledge and rely on the taught algorithms

(Hornburg et al., 2017). Furthermore, Laski et al. (2013) points out that girls' persistent counting provide them with fewer opportunities to practice derived fact strategies and retrieval and that might lead to girls' poorer accuracy in these strategies.

Considering the findings of this study, it would be relevant to investigate the effects of different aspects of instructional practices on sex specific development of strategies in arithmetic in the early years. For example, boys are known to be more willing to take risks in a competitive environment and it could be that teaching promoting a competitive learning environment may be more beneficial for boys, whereas girls might experience a reinforcement of the feedback loop (Bailey et al., 2012).

Limitations

We are aware of difficulties in researching with young children through one-to-one interviews (Punch, 2002), and that gender effects could be mediated by personality characteristics (Hornburg et al., 2017) such as impulsivity and inhibition which may affect strategies the student used in the interviews. Furthermore, using only a choice method, where the students are free to choose the strategy to solve each item, cannot necessarily provide information on the students' actual repertoire. There is a risk that the student's choice of strategy in the test situation is biased for different reasons as indicated above. To avoid such bias, a choice/no-choice method can be applied as discussed by Torbeyns, Verschaffel, and Ghesquière (2005). However, we believe this bias to be relatively small in this study, because of the students in general were used to this kind of test or interview situations in school, it took place in the students' familiar school settings and the students in most cases were familiar with the researcher.

The students' attempted category was not registered. As the error category was indirectly incorporated in the analysis of other strategy categories as each category was analysed against all other categories (including error) this could influence the result. However, as the proportional use of the error category was relatively small (less than 10% in year one) and decline rapidly with time the overall result would most likely not be influenced. Furthermore, Laski et al (2013) found that girls were least accurate when using counting strategies compared to direct retrieval and decomposition. If this is the case, then we would probably rather have underestimated the girls use of counting in year one.

The effect of sex in this study levelled out by year four in single-digit addition, as students move onto more challenging tasks by years two and three. This is not necessarily an indication that sex specific differences no longer are present. It is equally likely that it is caused by a ceiling effect similar to the ceiling effects reported by Shen et al. (2016). Thus, single-digit addition is probably not a good diagnostic tool for the average students beyond year three.

Conclusion and Perspectives

We have shown the existence of considerable specific sex differences in one of the foundational aspects of teaching number in the early years arithmetic, in the very first years of formal Danish schooling. Considering the lack of research on the learning and development of number and arithmetic in the early years of school in Denmark, as well as early sex differences, we find that it is important that these findings are reported.

The reasons behind these sex differences, equalling 2-3 year's, remain to be explicated. No matter the reasons behind these sex differences, their sheer magnitude in this as well as in at least one previous study (Bailey et al., 2012) may suggest that sex differences and their underlying causes deserves more attention than appear to be standard in most studies of arithmetic in early years of school.

Although a systematic review of the extent to which the presence and magnitude of sex differences in arithmetic research is beyond the scope of this paper, it seems that sex differences are sometimes not quantified in research papers, which means that sex differences in arithmetic patterns possibly could be underreported. The occurrence of differences in arithmetic patterns between boys and girls may be considered politically sensitive and vulnerable to partisan misinterpretations, but this does not change the fact that they sometimes exist and should be investigated, quantified and explained as any other predictor variable in education research. Knowledge on how individual differences and sociocultural factors interact to produce the resulting sex differences might thus prove useful to inform targeted teaching and educational practice that ‘decrease’ rather than unintentionally enhance sex differences no matter what have caused them.

References

Bailey, D. H., Littlefield, A., & Geary, D. C. (2012). The codevelopment of skill at and preference for use of retrieval-based processes for solving addition problems: Individual and sex differences from first to sixth grades. *Journal of Experimental Child Psychology, 113*(1), 78–92.

Benbow, C. P. (1988). Sex differences in mathematical reasoning ability in intellectually talented preadolescents: Their nature, effects, and possible causes. *Behavioral & Brain Sciences, 11,* 169–232.

Bremholm, J., & Skott, C. K. (2019). Teacher planning in a learning outcome perspective: A multiple case study of mathematics and L1 Danish teachers. *Acta Didactica Norge*, *13*(1), 1–22.

Browne, K. R. (2002). *Biology at work: Rethinking sexual equality.* New Brunswick, NJ: Rutgers University Press.

Canobi, K. H., Reeve, R. A., & Pattison, P. E. (1998). The role of conceptual understanding in children's addition problem solving. *Developmental Psychology, 34*, 882–891.

Carpenter, T. & Moser, J. (1984). The Acquisition of Addition and Subtraction Concepts in Grades One through Three. *Journal for Research in Mathematics Education*, *15*(3), 179–202.

Carr, M., & Alexeev, N. (2011). Fluency, accuracy, and gender predict developmental trajectories of arithmetic strategies. *Journal of Educational Psychology, 103*(3), 617–631. doi: 2048/10.1037/a0023864

Carr, M., & Davis, H. (2001). Gender differences in arithmetic strategy use: A function of skill and preference. *Contemporary Educational Psychology, 26*(3), 330–347. doi:10.1006/ceps.2000.1059

Carr, M., & Jessup, D. L. (1997). Gender differences in first-grade mathematics strategy use: Social and metacognitive influences. *Journal of Educational Psychology*, *89*(2), 318–328.

Carr, M., Steiner, H. H., Kyser, B., & Biddlecomb, B. (2008). A comparison of predictors of early emerging gender differences in mathematics competency. *Learning and Individual Differences, 18*(1), 61–75.

Casey, M. B., Nuttall, R., Pezaris, E., & Benbow, C. P. (1995). The influence of spatial ability on gender differences in mathematics college entrance test scores across diverse samples. *Developmental Psychology*, *31*(4), 697–705.

Connellan, J., Baron-Cohen, S., Wheelwright, S., Batki, A., & Ahluwalia, J. (2000). Sex differences in human neonatal social perception. *Infant Behavior & Development, 23,* 113–118.

Dowker, A. (2003). Young children’s estimates for addition: The zone of partial knowledge and understanding. In A. J. Baroody & A. Dowker (Eds.), *The development of arithmetic concepts and skills: Constructive adaptive expertise*. (pp. 243–266). Mahwah, NJ: Erlbaum.

Dowker, A. (2005). Early Intervention for students with mathematical difficulties. *Journal of Learning Disabilities, 38*(4) 324–332.

Dowker, A. (2009). Use of derived fact strategies by children with mathematical difficulties. *Cognitive Development, 24*(4), 401–410. doi:10.1016/j.cogdev.2009.09.005

Dowker, A. (2014). Young children's use of derived fact strategies for addition and subtraction. *Frontiers in Human Neuroscience, 8*, 1–9. doi:10.3389/fnhum.2013.00924

Else-Quest, N., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin, 136*(1), 103–127. http://dx.doi.org/10.1037/a0018053 . Retrieved 20.10.17

Fennema, E., Carpenter, T. P., Jacobs, V. R., Franke, M. L., & Levi, L. W. (1998). A longitudinal study of gender differences in young children's mathematical thinking. *Educational Researcher, 27*(5), 6–11.

Foley, A. E., Vasilyeva, M., & Laski, E. V. (2016). Children's use of decomposition strategies mediates the visuospatial memory and arithmetic accuracy relation. *British Journal of Developmental Psychology, 35*(2), 303–309. doi:10.1111/bjdp.12166

Geary, D. C. (2010). *Male, female: The evolution of human sex differences.* Washington, DC: American Psychological Association.

Geary, D. C., Bow-Thomas, C. C., Liu, F., & Siegler, R. S. (1996). Development of arithmetical competencies in Chinese and American children: Influence of age, language, and schooling. *Child Development, 67*, 2022–2044.

Geary, D. C., Hoard, M. K., & Nugent, L. (2012). Independent contributions of the central executive, intelligence, and in-class attentive behavior to developmental change in the strategies used to solve addition problems. *Journal of Experimental Child Psychology, 113*(1), 49–65. doi:10.1016/j.jecp.2012.03.003

Gersten, R., Jordan, N. C., & Flojo, J. R. (2005). Early identification and interventions for students with mathematics difficulties. *Journal of Learning Disabilities, 38*(4), 293–304.

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest*, *8*(1), 1–51.

Hirnstein, M., Andrews, L. C., & Hausmann, M. (2014). Gender-stereotyping and cognitive sex differences in mixed-and same-sex groups. *Archives of sexual behavior, 43*(8), 1663-1673.

Hornburg, C. B., Rieber, M. L., & McNeil, N. M. (2017). An integrative data analysis of gender differences in children's understanding of mathematical equivalence. *Journal of Experimental Child Psychology, 163*, 140–150.

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist, 60*(6), 581–592. doi:10.1037/0003-066X.60.6.581

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin, 107*(2), 139–155. doi: 10.1037/0033-2909.107.2.139

Hyde, J. S., & Mertz, J. E. (2009). Gender, culture, and mathematics performance. *Proceedings of the National Academy of Sciences, 106*(22), 8801–8807. doi:10.1073/pnas.0901265106

Imbo, I., & Vandierendonck, A. (2007). The development of strategy use in elementary school children: Working memory and individual differences. *Journal of Experimental Child Psychology, 96*(4), 284–309. doi:10.1016/j.jecp.2006.09.001

Laski, E. V., Casey, B. M., Yu, Q., Dulaney, A., Heyman, M., & Dearing, E. (2013). Spatial skills as a predictor of first grade girls' use of higher level arithmetic strategies. *Learning & Individual Differences, 23*, 123–130. doi:10.1016/j.lindif.2012.08.001

Laski, E. V., Ermakova, A., & Vasilyeva, M. (2014). Early use of decomposition for addition and its relation to base-10 knowledge. *Journal of Applied Developmental Psychology, 35*(5), 444–454.

Levine, S. C., Huttenlocher, J., Taylor, A., & Langrock, A. (1999). Early sex differences in spatial skill. *Developmental Psychology, 35*(4), 940–949.

Littell, R. C., Stroup, W. W., Milliken, G. A., Wolfinger, R. D., & Schabenberger, O. (2006). *SAS for mixed models.* SAS institute.

Maccoby, E. E., & Jacklin, C. N. (1974). *Psychology of sex differences.* Stanford, CA: Stanford University Press.

Moè, A. (2018). Mental rotation and mathematics: Gender-stereotyped beliefs and relationships in primary school children. *Learning and Individual Differences, 61*, 172–180.

Niss, M., & Højgaard, T. (2011). *Competencies and mathematical learning: Ideas and inspiration for the development of teaching and learning in Denmark* (IMFUFA tekst). Roskilde: Roskilde University.

Ostad, S. A. (1997). Developmental differences in addition strategies: A comparison of mathematically disabled and mathematically normal children. *British Journal of Educational Psychology, 67*(3), 345–357.

Paul, J. M., & Reeve, R. A. (2016). Relationship between single digit addition strategies and working memory reflects general reasoning sophistication. *Learning and Instruction, 42,* 113–122.

Penner, A. M., & Paret, M. (2008). Gender differences in mathematics achievement: Exploring the early grades and the extremes. *Social Science Research, 37*(1), 239–253. doi:http://dx.doi.org/10.1016/j.ssresearch.2007.06.012

Peterson, P. L., & Fennema, E. (1985). Effective teaching, student engagement in classroom activities, and sex-related differences in learning mathematics. *American Educational Research Journal*, *22*(3), 309–335.

Pinker, S. (2002). *The blank slate: The modern denial of human nature.* New York: Viking.

Pinxten, M., Marsh, H. W., De Fraine, B., Van Den Noortgate, W., & Van Damme, J. (2014). Enjoying mathematics or feeling competent in mathematics? Reciprocal effects on mathematics achievement and perceived math effort expenditure. *British Journal of Educational Psychology*, *84*(1), 152–174.

Price, G. R., Mazzocco, M. M., & Ansari, D. (2013). Why mental arithmetic counts: Brain activation during single digit arithmetic predicts high school math scores. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience, 33*(1), 156–163. doi:10.1523/JNEUROSCI.2936-12.2013

Punch, S. (2002) Interviewing strategies with young people: the 'secret box', stimulus material and task based activities. *Children and Society 16*(1): 45–56.

Royer, J. M., Tronsky, L. N., Chan, Y., Jackson, S. J., & Marchant, H. III, (1999). Math-fact retrieval as the cognitive mechanism underlying gender differences in math test performance. *Contemporary Educational Psychology*, *24*(3), 181–266.

Sarama, J., & Clements, D. H. (2009). *Early childhood mathematics education research: Learning trajectories for young children*. New York: Routledge.

Selter, C. (2001). Addition and subtraction of three-digit numbers: German elementary children's success, methods and strategies. *Educational Studies in Mathematics*, *47*(2), 145–173.

Shen, C., Vasilyeva, M., & Laski, E. V. (2016). Here, but not there: Cross-national variability of gender effects in arithmetic. *Journal of Experimental Child Psychology, 146*, 50–65. doi:10.1016/j.jecp.2016.01.016

Siegler, R. S. (1988). Individual differences in strategy choices: Good students, not-so-good students, and perfectionists. *Child Development, 59,* 833–851.

Skott, J. (2004). The forced autonomy of mathematics teachers. *Educational Studies in Mathematics*, *55*(1–3), 227–257.

Spelke, E. S. (2005). Sex differences in intrinsic aptitude for mathematics and science?: a critical review. *American Psychologist, 60*(9), 950–958.

Torbeyns, J., Verschaffel, L., & Ghesquière, P. (2005). Simple addition strategies in a first-grade class with multiple strategy instruction. *Cognition and Instruction, 23*(1), 1–21. Retrieved 21.4.18 from http://dx.doi.org/10.1207/s1532690xci2301_1

Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin, 140*(4), 1174–1204.

Table 1. Study A: Number of students assessed per class (Girls/Boys) at each year and mean dates for assessment interviews.

	Number of assessment interviews conducted per year										Students in total	
	Year 1		Year 2		Year 3		Year 4		Total			
Class	G/B	mean date	G/B	mean date	G/B	mean date	G/B	mean date	G/B	all	G/B	all
A1					11/6	2-Nov-13	13/6	30-May-15	24/12	36	13/6	19
A2			9/4	31-Aug-12	12/4	28-Oct-13	14/5	22-Apr-15	35/13	48	14/5	19
A3	0/1	1-Nov-12	11/6	29-Sep-13	14/8	21-Apr-15			25/15	40	14/8	22
A4	13/7	21-Feb-14	8/7	10-Apr-15	8/6	2-Jun-16			29/20	49	8/7	15
A5			7/10	1-Apr-15					7/10	17	7/10	17
A6	6/10	26-Apr-15	6/9	6-Jun-16					12/19	31	6/10	16
A7	6/9	21-Apr-15							6/9	15	6/9	15
A8	9/15	6-Oct-15							9/15	24	9/15	24
Sum	34/42		41/36		45/24		27/11		147/113	260	77/70	147

Table 2. Study B: Number and mean dates of assessment interviews conducted on students (Girls/Boys).

	Number and date of assessment interviews:						Students in total	
	round A		round B		Total			
Class	G/B	mean date	G/B	mean date	G/B	all	G/B	all
B1	9/6	5-Nov-15	8/5	14-Apr-16	17/11	28	9/6	15
B2	5/5	3-Nov-15	7/5	20-Apr-16	12/10	22	6/5	11
B3	7/5	2-Nov-15	7/5	14-Apr-16	14/10	24	7/5	12
B4	6/5	10-Nov-15	7/6	15-Mar-16	13/11	24	8/6	14
B5	7/6	5-Nov-15	6/6	4-Apr-16	13/12	25	7/6	13
B6	9/9	26-Oct-15	8/6	29-Mar-16	17/15	32	9/9	18
Total:	43/36	2-Nov-15	43/33	5-Apr-16	86/69	155	46/37	83

Table 3. Equations for GLMMs (logit link) of proportional use of the four different addition strategies in study A as interactive functions of year (range from mid year 1 to mid year 4) and sex, when accounting for random variation attributable to the individual, student specific assessment situation, student ID, class and addition task. Statistical significances (unless otherwise stated F- and t-statistics gave similar results): ns: $p \geq 0.1$, °: $p < 0.1$,*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$, ****: $p < 0.0001$.

	Error					Counting				
Fixed effects	B	SE(B)	DF	t/z	Sign	B	SE(B)	DF	t/z	Sign
Intercept	-0.86	0.63	80	-1.35	ns	-0.39	0.32	57	-1.21	ns
Sex (S)	-0.40	0.71	199	-0.55	ns	1.31	0.33	254	3.99	****
Year (YR)	-1.63	0.26	178	-6.28	****	-0.36	0.09	207	-4.06	****
YR*S	0.45	0.30	235	1.50	ns	-0.17	0.10	227	-1.58	ns
Random effects (covariance parameters):										
Test(student)	0.62	0.27		2.27	*	0.38	0.07		5.41	****
student(class)	1.75	0.43		4.09	****	0.64	0.13		5.00	****
class	0.18	0.22		0.82	ns	0.21	0.13		1.53	ns
Addition task	1.39	0.42		3.30	***	0.45	0.11		4.03	****
	Direct retrieval					Derived fact				
Fixed effects	B	SE(B)	DF	t/z	Sign	B	SE(B)	DF	t/z	Sign
Intercept	-1.05	0.26	65	-3.99	***	-2.17	0.58	79	-3.76	***
Sex (S)	-0.74	0.19	254	-3.88	****	-1.64	0.53	210	-3.07	**
Year (YR)	0.24	0.05	197	5.29	****	0.26	0.14	139	1.82	°§
YR*S	0.12	0.06	197	1.99	*	0.30	0.17	183	1.74	°
Random effects (covariance parameters):										
Test(student)	0.06	0.02		2.56	*	1.43	0.24		5.96	****
student(class)	0.23	0.04		5.41	****	0.94	0.29		3.23	**
class	not estimable					0.30	0.24		1.27	ns
Addition task	1.82	0.43		4.20	****	4.45	1.14		3.89	***

§ in this model, the t-statistics for year yielded a p-value of 0.08, whereas F-statistics resulted in $p < 0.0001$. If the interaction term was removed from the model, the t- as well as the F-statistics suggested a strongly significant effect ($p < 0.0001$) of year.

Table 4. Comparison and tests for differences between study A and B in predicted relative differences of strategy use by girls and boys (logit scale) in mid year one as well as the mean probabilities (also on logit scale) by which girls and boys were using the different strategies to solve simple addition tasks. Statistical significances: ns: $p \geq 0.1$, °: $p < 0.1$,*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$, ****: $p < 0.0001$.

		Sex difference (G-B)					Mean: Girls					Mean: Boys				
Strategy	Study	B	se	df	t	sign	b	se	df	t	sign	b	se	df	t	sign
error	A	0.14	0.43	189.5	0.34	ns	-3.03	0.36	34.7			-3.31	0.38	31.3		
	B	0.49	0.36	84.9	1.37	ns	-4.01	0.39	15.5			-4.50	0.42	20.8		
	Diff A-B	-0.35	0.56	274.4	-0.62	ns	0.99	0.53	50.2	1.85	°	1.19	0.56	52.1	2.11	*
counting	A	1.11	0.23	241.4	4.83	****	0.12	0.24	23.7			-0.94	0.25	24.8		
	B	1.37	0.31	74.5	4.50	****	0.61	0.31	76.6			-0.76	0.32	85.0		
	Diff A-B	-0.27	0.38	315.9	-0.70	ns	-0.49	0.39	100.3	-1.25	ns	-0.18	0.41	109.8	-0.43	ns
direct retrieval	A	-0.60	0.13	263.9	-4.43	****	-1.25	0.24	47.1			-0.68	0.24	46.3		
	B	-0.90	0.26	73.7	-3.45	****	-1.16	0.25	45.7			-0.66	0.25	47.8		
	Diff A-B	0.30	0.29	337.6	1.03	ns	-0.09	0.35	92.8	-0.25	ns	-0.03	0.35	94.1	-0.08	ns
derived fact	A	-1.29	0.37	217.1	-3.52	***	-2.98	0.47	49.3			-1.78	0.47	46.7		
	B	-1.49	0.37	65.0	-3.97	***	-2.44	0.24	38.0			-1.24	0.24	40.9		
	Diff A-B	0.20	0.52	282.1	0.39	ns	-0.54	0.53	87.3	-1.02	ns	-0.54	0.53	87.6	-1.03	ns

Table 5. Test statistics for between-class effects (6 classes) in study B in statistical models where class identity and sex-by-class identity were entered as fixed effects. The models also accounted for random effects of test case nested within student identity, student identity and addition task identity. For simplicity these results are not presented. Statistical significances: ns: $p \geq 0.1$, *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$, ****: $p < 0.0001$.

	Error			Counting			Direct retrieval			Derived fact		
Fixed effect	F	df	Sign	F	df	Sign	F	df	Sign	F	df	Sign
Sex	0.51	1,70.2	ns	19.64	1,67.2	****	10.34	1,65.1	**	13.1	1,57.1	***
Year	19.13	1,66.5	****	17.12	1,68.7	****	22.63	1,63.9	****	19.31	1,50.0	****
classID	2.62	5,68.4	*	0.7	5,67.4	ns	0.62	5,64.8	ns	1.47	5,57.2	ns
Sex*classID	0.65	5,68.2	ns	0.33	5,67.3	ns	0.37	5,64.7	ns	0.71	5,57.1	ns

Figure 1: Proportional use of different addition strategies plotted against school year for girls and boys. Thick lines indicate the predicted functions for each sex with thin lines demarcating 95% confidence zones.

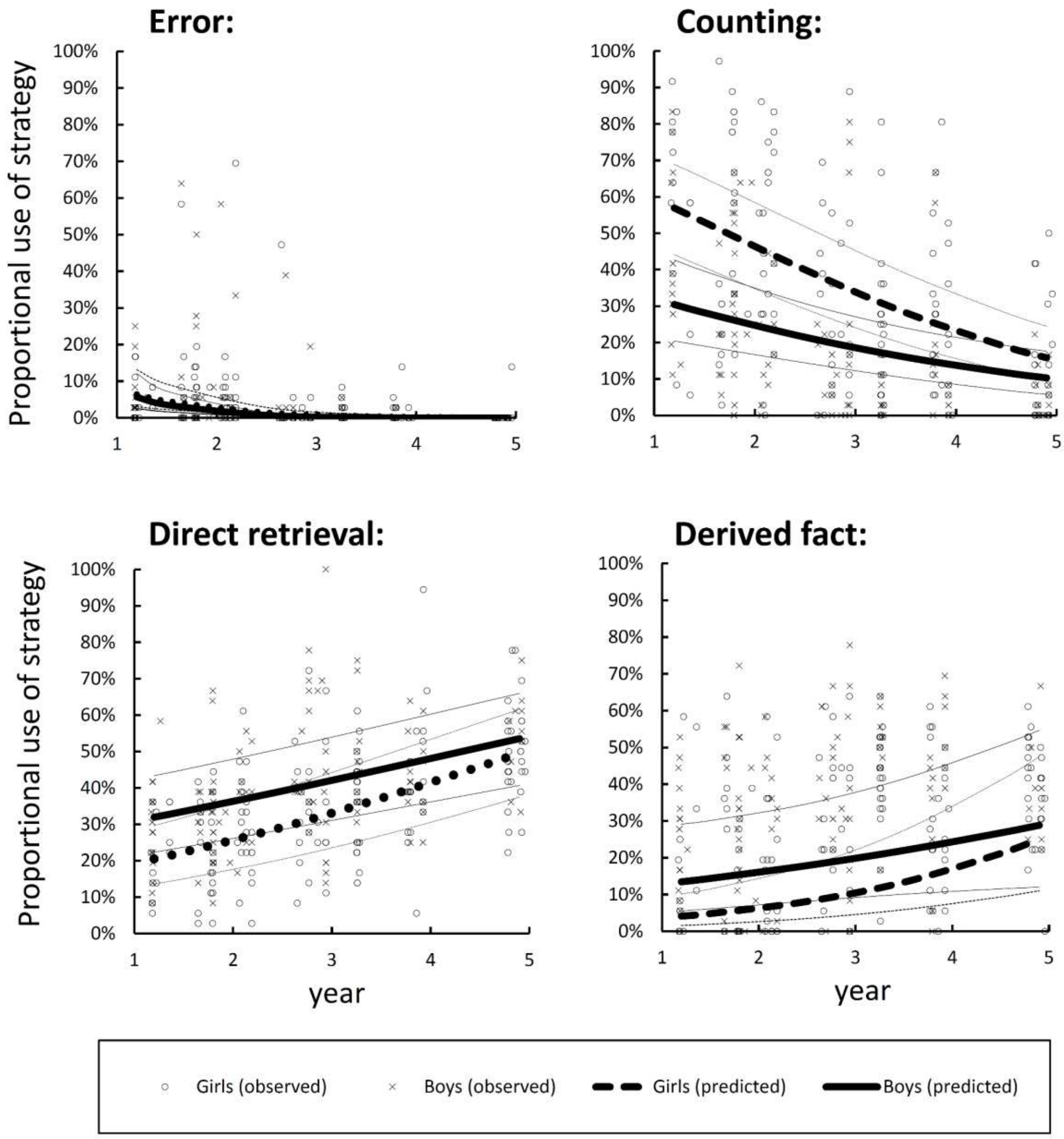

Figure 2. Strategy use by mid year one (least square means that account for random variation: error bars indicate 95% confidence intervals) by girls and boys divided on study and class within study B.