Title
Intuitive Statistics: Identifying Children’s Data Comparison Strategies using Eye Tracking

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 36(36)

ISSN
1069-7977

Authors
Morris, Bradley
Cravalho, Patrick
Junglen, Angela
et al.

Publication Date
2014

Peer reviewed

Intuitive Statistics: Identifying Children's Data Comparison Strategies using Eye Tracking

Bradley J. Morris (bmorri20@kent.edu), Patrick F. Cravalho (pcravalh@kent.edu), Angela Junglen (ajungle1@kent.edu), Christopher Was (cwas@kent.edu)
Department of Educational Psychology, 405 White Hall
Kent, OH 44242 USA

Amy M. Masnick (Amy.M.Masnick@hofstra.edu)
Department of Psychology, Hauser Hall
Hempstead, NY 11549 USA

Abstract

People often compare sets of numbers informally, in considering prices or sports performance. Children who lack knowledge of formal comparison strategies (e.g., statistics) may use intuitive strategies like estimation that create summary values with approximations of means and variance. There were two goals for this experiment: (1) to classify data comparison strategies and (2) to evaluate whether children's strategy discovery and selection is effective. Using eye tracking, we identified strategies used by 41 8-12-year-old children when comparing number sets, by examining how the properties of the data sets (e.g., mean ratios and variance) influenced accuracy and confidence in differences. We classified strategies from eye tracking patterns; these strategies were associated with different levels of accuracy, and strategy selection was adaptive in that selection was related to the statistical properties of the sets being compared. The results demonstrate that children are quite adept at informally comparing data sets and adaptively select strategies to match the properties of the sets themselves.

Keywords: Intuitive statistics, cognitive development, eye tracking

Introduction

How does a shopper determine which store has the lowest prices? This could be achieved by comparing sets of prices using a formal statistical analysis like a t-test. However, when comparing numbers in context (data), outside of a research setting, such comparisons are more likely to occur informally. Surprisingly, there has been relatively little research investigating how people compare sets of numbers (e.g., Morris & Masnick, in press) despite the rapidly growing literature investigating comparisons between single numbers (e.g., Dehaene, 2009). In the current paper, we investigate the comparison of number sets in children using eye tracking in order to classify and evaluate naïve strategies for performing data comparisons.

Number representation and comparison

Numbers are represented as both a verbal category (e.g. "twelve", an exact value) and an activation function on an approximate number system (Dehaene, 2009; Feigenson, Dehaene, & Spelke, 2004; Opfer & Siegler, 2012). Differences between single-digit numbers are detected more quickly and accurately as the ratio of the numbers increases (Dehaene, 2009). For example, reaction times are slower and evaluations are less accurate when comparing 9 and 10 (9:10 ratio of numbers), than when comparing 3 and 9 (1:3 ratio of numbers). This distance effect is evidence that numerical quantities are compared using approximate representations. This effect is detected across species (Brannon, 2003), across age of participants (Feigenson et al., 2004), and across presentation formats (e.g., dots, Arabic numbers, fractions; Buckley & Gillman, 1974; Dehaene, 2001; Sprute & Temple, 2011).

Multi-digit number comparisons (e.g., 63 vs. 72) demonstrate a unique distance effect in that two- and three-digit numbers produce distance effects for each place value unit (i.e., tens vs. ones; Korvorst & Damian, 2008; Nuerk, Weger, & Willmes, 2001). Eye tracking results demonstrate more fixations when there is incompatibility between places (e.g., 47 vs. 51), where one number has a larger value in the tens column, but the other one has a larger value in the ones column (Moeller, Fischer, Nuerk, & Willmes, 2009).

The Approximate Number System (ANS) allows children, even infants, to discriminate quantities, although the threshold at which accurate discrimination occurs changes over development (Mou & vanMarle, in press). Specifically, infants can distinguish quantities at a ratio of 1:2; by early elementary school, children distinguish at a 4:5 ratio; and adults discriminate quantities at a 9:10 ratio (Halberda & Feigenson, 2008; Xu & Spelke, 2000). Children are also adept at using the ANS to estimate solutions to problems for which they lack formal solution strategies (Gilmore, McCarthy, & Spelke, 2007). One explanation for the increasing acuity is the strategies children (and adults) use to compare quantities (Hyde, 2011; Mu & van Marle, in press). That is, strategies are related to goals (e.g., accuracy) and constrained by processing limitations (e.g., working memory; Hyde, 2011).

Comparing number sets

There is evidence that when comparing number sets, adults create and compare approximate summaries of the statistical properties of these sets, which include information about means and variances (Morris & Masnick, in press; Masnick & Morris, 2008; Obrecht, Chapman, & Suarez, 2010). For example, adults were asked to compare sets of 3-digit numbers that varied in set size (4 or 8), ratio of means (2:3, 4:5, or 9:10), and coefficient of variation (.10 or .20 of mean). The results suggested that summaries were compared much like single values in that accuracy and confidence were higher and visual fixations shorter for comparisons between high contrast sets (e.g., high mean ratio, low variance) than low contrast sets (e.g., low mean ratio, high variance; Morris & Masnick, in press). Children as young as 8 show some perception of variance in data sets, reducing their confidence in conclusions with more varied data, though they do not change their confidence as much as adults (Masnick & Klahr, 2003; Masnick & Morris, 2008).

One model suggests that summaries emerge from multiple activation areas in the approximate number system (Morris & Masnick, in press). More specifically, approximate means emerge from the overlap between two or more values and approximate variance could emerge from spread of values within the set. Sets with large mean ratios and low variance would constitute high contrast sets, in that they would create summary activation areas that would be far apart on a number line (low contrast sets would be close together).

In order to create the aforementioned summary representations, a reasoner must use a strategy that encodes information necessary to derive summary values. For example, given two columns of three-digit numbers, when an adult attends to the hundreds place value of both columns, she can quickly detect differences given high contrast sets (e.g., left has mostly 4s and 5s in the hundreds column, while the right has mostly 7s and 8s, therefore the right column is larger). However, if a second reasoner only attends to the first three-digit numbers in each set, then he will not encode the properties of the set, which will influence the types of possible comparisons. Adults are consistent in attending to the numeric properties of an entire data set (Masnick & Morris, 2008; Morris & Masnick, in press).

Children are often less consistent than adults, demonstrating high variability in strategy use (Siegler, 2007). Although more variable, children often demonstrate highly adaptive strategy use in that the selection of strategies is related to processing goals (Siegler, 1996). When solving complex addition problems, children discover new strategies without conscious awareness and modify their strategies to improve solution accuracy without explicit feedback (Siegler & Stern, 1998). Thus, we investigated naïve data comparison strategies of elementary school children.

In the current study, we asked children to compare data presented as columns of three-digit numbers (framed as the distances two golfers drove a golf ball, when doing so repeatedly), and to choose which golfer hit the ball farther. We examined potential strategies for assessing such differences between data sets, and then linked these strategies to specific behavioral predictions. Table 1 summarizes the set of possible strategies and predicted behavior, given each strategy, described in more detail below. We suggested an additional strategy in which participants sought the highest or lowest value. However, no subject used this strategy and it was eliminated from our coding.

Data Comparison Strategies

One possible approach for comparing sets of numbers is to attend to only a subset of the number set. We suggest two possible types of subset strategies. The use of any of these strategies reduces information about set properties.

Strategy 1: Pairwise comparisons. One possible subset strategy is to make paired comparisons, comparing individual values within one set to the individual values within the other set. This strategy would involve direct comparisons within each pair of data, such as noting which golfer had the longer drive on each trial. After comparing each pair of drives, the Golfer with the most "wins" (e.g., 5 of 6 times one golfer had a longer drive) would be determined to have hit the ball farther. From a processing standpoint, such a strategy would require a series of comparisons that determine the larger value and a simple tally of the results. Even if used on all pairs in a data set, such comparisons would ignore potentially relevant information about the statistical properties of the set that may influence the overall evaluation (e.g., mean ratios, variance).

Strategy 2. Subset comparisons. A second strategy is comparing a sample of values, such as comparing only the first two drives in a set and ignoring the remaining numbers. Using this strategy, comparisons for sets would operate exactly as predicted in the comparison between two numbers. From a processing standpoint, such a strategy would not require any calculation.

Strategy 3: Calculation. Another possible strategy for comparing sets is for reasoners to calculate values explicitly. For example, one might compute a mean value for each column and then compare these means. One might also add the values and compare the sums (which would lead to the same conclusion as computing means).

Strategy 4: Gist. Another strategy for comparing sets of numbers is to summarize the statistical properties of the numbers, including unique characteristics of sets unavailable in individual number comparisons, such as means and variances. This strategy would lead to a quick decision, but one that was based on a rough estimation of mean and spread of each data set. Such summaries likely emerge from scanning a set of numbers because the activation of multiple approximate magnitudes for each member of the set results in a summary activation area in regards to a number line. Sets with large mean ratios and low variance would constitute high contrast sets, in that they would create summary activation areas that would be far

Table 1. Data comparison strategies, descriptions, and hypothesized response patterns

Strategy	Accuracy	Confidence	Eye fixation pattern
Pairwise Comparison	High with high contrast sets.	No change with sample size.	Sequential pattern of fixations comparing pairs within each row (e.g., fixations targeted between first two numbers in sets) before moving on to next set until all pairs have been compared.
Subset Comparison	High with high contrast sets	No change with sample size	Fixations directed to only a subset of pairs (e.g. first two, last two). No fixations on more than half of all numbers in the sets.
Calculation	High; No change with set properties.	Decreases as sample size increases	Long fixations on place values or entire numbers in each set
Gist	Increases as set contrast increases.	Increases with high contrast sets, smaller sample sizes.	Rapid scanning by place values of numbers; often associated with decomposition of numbers (e.g., focus on hundreds place value only)
Gist +	Increases as set contrast increases.	Increases with high contrast sets, smaller sample sizes.	Initial scan strategy with additional information pickup. For example, subjects may iteratively scan different place values (e.g., tens) or whole number comparisons.

apart on a number line. If the summary representations are compared like single number representations, then a distance effect for sets would be expected, and high contrast sets would be associated with faster, more accurate comparisons and low contrast sets with slower, less accurate comparisons.

Strategy 5: Gist+. This strategy begins as the Gist strategy described above. However, instead of terminating the search and producing a response, subjects continue to search for additional information, update information in working memory, or process information. One key difference from the Gist strategy is that, without detecting a clear difference between the two sets, subjects may simply continue to scan for more information. In some cases, this scanning might be deliberate. For example, subjects may scan, then conduct pairwise comparisons between whole numbers in the sets or may search for the highest/lowest value before responding. In other cases, the scanning may be incidental. For example, additional scanning might not be directed towards specific elements (e.g., iterative scans of tens columns), but might be simply looking at various elements until they reach either some criterion or simply terminate the search and produce a guess.

Current paper. We investigated how children compare number sets by presenting participants with sets of data (i.e., numbers in context) that varied systematically in the mean ratio, relative variance, and number of observations. We asked participants to compare these sets, and we measured eye fixations, accuracy, and confidence in comparisons to investigate set comparisons when the mean, variance and number of observations were systematically manipulated. From these data, we classified their comparison strategies, evaluated the accuracy of strategies, and investigated whether children's strategy discovery and selection was adaptive.

Method

Participants

Participants were 41 8-12 year-old children (M = 10.76, SD = .46) who returned a signed parental consent form and provided verbal assent to participate. One participant was not included in the analysis because of problems with the eye-tracking recording.

Procedure

Participants saw 36 data set pairs with the following properties: (a) set size 4, 6, or 8, (b) low (9:10) or high (4:5) ratio of means, and (c) either low or high variance. Numbers were presented in 42-point Times New Roman font and each column of numbers was centered within two columns in a table. Within each number, an extra space was placed between the hundreds, tens, and ones values and 1.5 spacing was used between numbers in each column.

Participants were asked to determine which golfer (left or right) hit the ball farther (accuracy) and rate their confidence in this evaluation on a 4-point scale (1 = Not at all sure, 4 = Totally sure). On each trial, participants first saw a fixation slide in which a + was placed in the center of the screen for 1 second. After one second, participants saw a data slide consisting of the two sets of data (positioned on the left or right side of the screen) and were given unlimited time to view the datasets. Participants were then asked to determine which golfer, on average, hit the ball farther and how confident they were in this difference (using the scale positioned in front of them). Data sets were presented sequentially in blocks by sample size (4, 6, or 8 observations) because pilot testing demonstrated that the change in stimuli between presentations (e.g., presenting a set of four, then a set of eight) resulted in a higher number of fixations compared to blocked presentations. In this way, blocked presentation controls for errant fixations and focuses attention on features associated with the stimuli, not with the mode of presentation.

Data sets were presented on a Tobii® T-60XL eye tracker. Participants were given a 9-point calibration before beginning the task. Areas of Interest (AOIs) were recorded around the hundreds, tens, and ones columns and around whole numbers. Inter-rater reliability was 92% before discussion, and all disagreements were resolved with the first author.

Results

Aggregated Results

Accuracy and Confidence. A 3-way repeated measures ANOVA was run in which mean ratio (4:5; 9:10), coefficient of variation (low, high) and sample size (4, 6, 8) were independent variables, and accuracy (out of 3) was the dependent measure. For accuracy, there was a main effect for mean ratio, with lower accuracy for sets with 9:10 ratio ($M = 2.33$, $SD = .4$), and higher accuracy for 4:5 sets (M = Figure 1). There were no significant differences for confidence ratings.

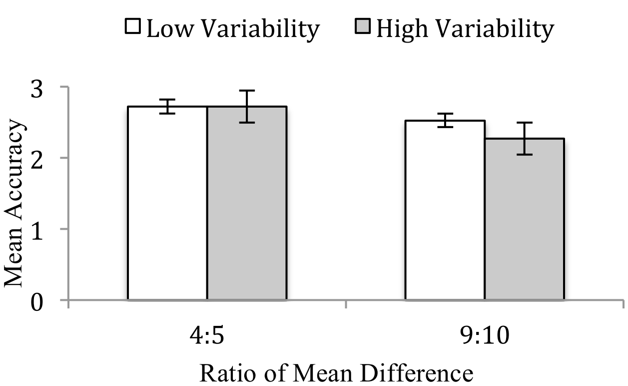

Figure 1. Mean accuracy by condition.
Note. Error bars represent SE of Mean.

Eye Fixation Counts. Eye fixation counts were used to explore the pattern of where people look for information in this task. There were main effects for all variables in our 4-way ANOVA in which mean ratio (4:5, 9:10), coefficient of variation (low, high), sample size (4, 6, 8), and column (hundreds, tens, ones) were all repeated measures independent variables, and number of fixations with the Areas of Interest defined by each of the six columns (hundreds, tens, ones on the left and right sides of the screen) was the dependent measure. A repeated measure ANOVA demonstrated two main effects: a significant overall increase in the number of fixations as set size increased, $F(2, 39) = 6.6$, $p = .01$, $\eta2 = .32$, and significantly more fixations in the hundreds column than in the tens or one column, $F(2, 39) = 21.8$, $p = .001$, $\eta2 = .61$. There were more fixations in the hundreds column as the mean difference between sets decreased $F(2, 39) = 14$, $p = .001$, $\eta2 = .45$; more fixations in the hundreds column as variance increased $F(2, 39) = 8.8$, $p = .01$, $\eta2 = .39$; and significantly more fixations in the hundreds column for low contrast compared to high contrast sets $F(3, 39) = 12$, $p = .001$, $\eta2 = .36$. An additional ANOVA was run with fixation duration as the dependent measure, and it yielded a similar pattern of results.

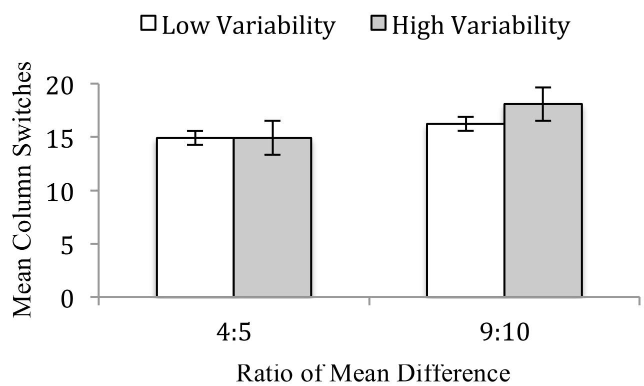

Figure 2. Mean number of column switches by condition.
Note. Error bars represent SE of Mean.

Column switches. Another measure of comparison was the number of times people moved their eyes from one side to the other. In reading comprehension research, regressions are measured as fixations in which subjects re-read text, and are an accurate index of text complexity (Rayner & Pollatsek, 1989). Here, we considered looking at a column of data repeatedly in a similar manner. Two raters counted the number of times subjects switched fixations between columns. The two coders agreed 95% of the time and discrepancies were resolved through discussions with the first author. We then examined the effect of data characteristics on the frequency of fixation switches. A mean ratio (4:5, 9:10) x coefficient of variation (low or high) x sample size (4, 6, 8) repeated measures ANOVA was conducted with the number of fixation switches as the dependent measure. Overall, there were more column switches given low mean ratios ($F(1, 39) = 53.3$, $p < .01$, partial $\eta2 = .44$), and high variance ($F(1, 39) = 36$, $p < .01$, partial $\eta2 = .36$). There was also an interaction in which switches were more frequent with low contrast sets ($F(2, 39) = 16.6$, $p < .01$, partial $\eta2 = .28$; See Figure 2).

Strategy analysis

In order to investigate data comparison strategies, we examined the patterns of eye motions to compare them to the proposed patterns associated with strategies outlined in Table 1. Two raters coded the eye movement patterns. They had 87% agreement, and resolved discrepancies after discussion with the first author.

Analysis showed substantial variability across the set of trials. Ninety-five percent of participants used at least 3 strategies during the experiment. The pairwise comparison and calculation strategies were excluded from further analyses because one participant used them each once. A McNemar Chi-square was used to compare strategy use between high and low contrast sets only. The use of the Gist+ strategy was significantly higher for low contrast sets while the use of the Gist strategy was significantly higher for high contrast sets (40, df = 6, p = .002; See Figure 3).

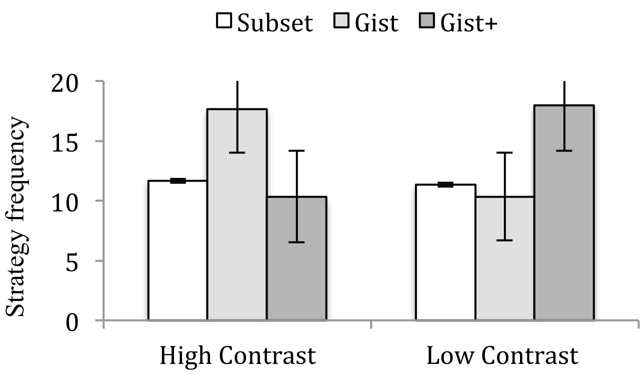

Figure 3. Strategy frequency by condition. Note. Error bars represent SE of Mean

A mean ratio (4:5, 9:10) x coefficient of variation low or high) x sample size (4, 6, 8) repeated measures ANOVA was conducted comparing accuracy by strategy codes. The results indicate that there was no difference in strategy accuracy for high contrast sets ($F\ (1, 39) = 1.2$, $p > .10$). For low contrast sets, there was a significant relationship between strategy and accuracy. Specifically, the comparison strategy was associated with significantly lower accuracy than the Gist or Gist+ strategies ($F\ (1, 39) = 24.5$, $p < .01$, partial $\eta 2 = .23$) and the Gist+ strategy was associated with lower accuracy than the Gist strategy ($F\ (2, 39) = 13.9$, $p < .01$, partial $\eta 2 = .15$; See Figure 4). Unlike accuracy, confidence ratings did not significantly differ by strategy or sample size. The subset strategy was used more frequently for high contrast sets than for low contrast sets. The subset strategy was as accurate for high contrast sets as Gist and Gist+ but was significantly less accurate for low contrast sets $F\ (2, 39) = 21$, $p < .001$, partial $\eta 2 = .30$.

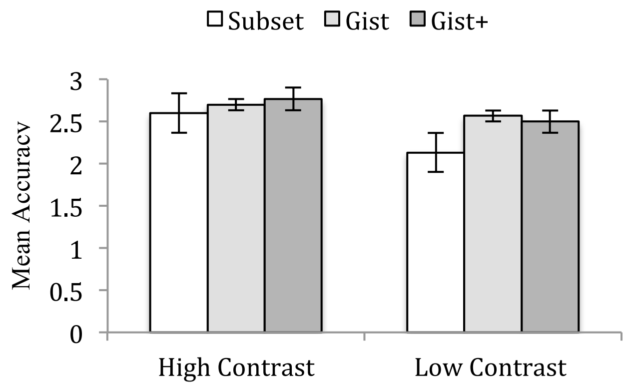

Figure 4. Strategy accuracy by condition. Note. Error bars represent SE of Mean

Discussion

Our results provide insights into how children make sense of data. Recall that children were asked to compare a series of data sets that differed in their statistical properties (e.g., mean ratio, variance). Adults are quite adept at such comparisons (Morris & Masnick, in press); however, adults likely have greater knowledge of number properties and mathematical operations as well as greater experience with such comparisons than children.

Our results provide four main findings. One, children were highly accurate at comparing data sets even when those sets were quite similar (e.g., 9:10 ratio of mean difference with high variability). Most children attended to highly relevant information when making comparisons. For example, children were more likely to attend to the information in the hundreds place value when making comparisons.

Two, children used a variety of strategies to make these comparisons. For example, nearly all children used at least three different strategies during the experiment. Three, changes in strategy use were associated with changes in the statistical properties of the data sets. For example, the use of a Gist strategy was more frequent with high contrast than with low contrast sets and the use of the Gist+ strategy was more frequent with low contrast set than high contrast sets. This change may indicate a purposeful search for additional information (e.g., tens column), re-activation of information in working memory that was previously scanned (e.g., fixating on hundreds column for a second time), or simply might indicate additional looking time associated with a lack of certainty regarding any conclusion.

Finally, strategies were associated with accuracy. For high contrast sets, there was no difference in accuracy; however, strategy use was associated with accuracy for low contrast sets. Specifically, the use of the subset strategy, in which children attended to the first two numbers in the set, was significantly less accurate than strategies in which children attended to the entire set. For high contrast sets it is highly probable that comparing any two individual values in the set would be diagnostic of the overall differences between sets. For low contrast sets, this strategy was significantly less accurate because the comparison of two values from the set are much less likely to be diagnostic of set differences than in high contrast sets. The use of Gist and Gist+ strategies would provide information about the entire set, which would provide information about the statistical properties of the set. Interestingly, for low contrast sets the additional information gained from the Gist+ strategy did not provide a significant increase in accuracy above the Gist strategy. It is possible that children either focused on irrelevant information in these additional fixations or children exceeded their working memory capacity with the additional information.

In sum, many 8-12-year-old children displayed an intuitive understanding of the statistical properties of number sets, using them to draw conclusions and evaluate strategy application. This nascent understanding demonstrates that the Approximate Number System may support rapid, implicit comparisons between number sets without the need for deliberate calculation. These findings may be able to inform formal scientific and statistical training by highlighting critical elements in creating effective problem representations and developing effective instructional tools.

References

Brannon, E. M. (2003). Number knows no bounds. *Trends in Cognitive Sciences, 7(7)*, doi:10.1016/S1364-6613(03)00137-2

Buckley, P. B., & Gillman, C. B. (1974). Comparisons of digits and dot patterns. *Journal of Experimental Psychology, 103(6)*, 1131-1136. doi:10.1037/h0037361

Dehaene, S. (2009). Origins of Mathematical Intuitions: The Case of Arithmetic. *The Year in Cognitive Neuroscience 2009: Annual New York Academy of Science, 1156*, 232–259. doi: 10.1111/j.1749-6632.2009.04469

Dehaene, S. (2001). Précis of the number sense. *Mind & Language, 16*, 16-36.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences, 8(7)*, 307-314. doi:10.1016/j.tics.2004.05.002

Gilmore, C. K., McCarthy, S. E., & Spelke, E. S. (2007). Symbolic arithmetic knowledge without instruction. *Nature, 447(7144)*, 589-591. doi:10.1038/nature05850

Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the “Number Sense": The Approximate Number System in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental psychology*, *44*(5), 1457.

Hyde, D. C. (2011). Two systems of non-symbolic numerical cognition. *Frontiers in human neuroscience*, *5, 150,* 1-8.

Korvorst, M., & Damian, M. F. (2008). The differential influence of decades and units on multidigit number comparison. *The Quarterly Journal of Experimental Psychology, 61(8)*, 1250-1264. doi:10.1080/17470210701503286

Masnick, A. M., & Klahr, D. (2003). Error Matters: An initial exploration of elementary school children's understanding of experimental error. *Journal of Cognition and Development, 4(1)*, 67-98.

Masnick, A. M., & Morris, B. J. (2008). Investigating the development of data evaluation: The role of data characteristics. *Child Development, 79(4)*, 1032-1048. doi:10.1111/j.1467-8624.2008.01174.x

Moeller, K., Fischer, M. H., Nuerk, H., & Willmes, K. (2009). Sequential or parallel decomposed processing of two-digit numbers? Evidence from eye-tracking. *The Quarterly Journal of Experimental Psychology, 62(2)*, 323-334. doi:10.1080/17470210801946740

Morris, B. J., & Masnick, A. M. (in press). Comparing data sets: Implicit summaries of the statistical properties of number sets. *Cognitive Science: A Multidisciplinary Journal.*

Mou, Y., & vanMarle, K. (in press). Two core systems of numerical representations in infants. *Developmental Review.*

Nuerk, H., Weger, U., & Willmes, K. (2001). Decade breaks in the mental number line? Putting the tens and units back in different bins. *Cognition, 82(1)*, B25-B33. doi:10.1016/S0010-0277(01)00142-1

Obrecht, N. A., Chapman, G. B., & Suárez, M. T. (2010). Laypeople do use sample variance: The effect of embedding data in a variance-implying story. *Thinking & Reasoning, 16(1)*, 26-44. doi:10.1080/13546780903416775

Opfer, J. E., & Siegler, R. S. (2012). Development of quantitative thinking. In K. Holyoak and R. Morrison (Eds.), *Oxford Handbook of Thinking and Reasoning*.

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Englewood Cliffs, NJ US: Prentice-Hall, Inc.

Siegler, R. S. (2007). Cognitive variability. *Developmental Science, 10(1)*, 104-109.

Siegler, R. S. (1996). *Emerging minds: The process of change in children’s thinking.* New York: Oxford University Press.

Siegler, R. S., & Stern, E. (1998). Conscious and unconscious strategy discoveries: a microgenetic analysis. *Journal of Experimental Psychology: General, 127(4)*, 377.

Sprute, L., & Temple, E. (2011). Representations of fractions: Evidence for accessing the whole magnitude in adults. *Mind, Brain, and Education, 5(1)*, 42-47. doi:10.1111/j.1751-228X.2011.01109.x

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition, 74(1)*, B1-B11.