Examining and Developing Fourth Grade Children's Area Estimation Performance

Submitted Draft Version 2018.

Authors: C. Eames, J. Barrett, C. Cullen, G. Rutherford, D. Klanderman, D. Clements, J. Sarama, & D. VanDine.

Appears in *School Science and Mathematics* (2020). Volume 120, Pages 67-78.

The research reported here was supported by the National Science Foundation under Grant No. DRL-1222944. The findings and statements in this manuscript do not represent the views of the National Science Foundation. A subset of the results discussed in this manuscript were presented at the 2015 Annual Meeting of the American Educational Research Association and the 2016 Research Session of the Annual Meeting of the National Council of Teachers of Mathematics.

Examining and Developing Fourth Grade Children's Area Estimation Performance

This study explored children's area estimation performance. Two groups of fourth grade children completed area estimation tasks with rectangles ranging from 5 to 200 square units. A randomly assigned treatment group completed instructional sessions that involved a conceptual area measurement strategy along with numerical feedback. Children tended to underestimate areas of rectangles. Furthermore, rectangle size was related to performance such that estimation error and variability increased as rectangle size increased. The treatment group exhibited significantly improved area estimation performance in terms of accuracy, as well as reduced *variability and instances of extreme responses. Area measurement estimation findings are related to a Hypothetical Learning Trajectory for area measurement.*

The pedagogical affordance of estimation as a means for assessing the reasonableness of a calculation or measurement has been repeatedly emphasized by standards documents in many countries, such as Japan (Takahashi, Watanabe, & Yoshida, 2008), Australia (Australian Curriculum, Assessment, and Reporting Authority [ACARA], 2014), the United Kingdom (Department for Education, 2013), and the United States (National Governors Association Center for Best Practices and Council of Chief State School Officers [NGA & CCSSO], 2010; National Council of Teachers of Mathematics [NCTM], 1989, 2000). The curricular guidelines and recommendations from countries as diverse as China, India, and Japan have consistent emphases on teaching both estimation and measurement for length and area topics as early as Year 1 and extending on through Year 6 (Cheung Kwok-cheung, 2004; Members of National Focus Group on Teaching of Mathematics [India], 2006; Takahashi, Watanabe, & Yoshida [Japan], 2008). The National Council of Educational Research and Training for India argues that mathematical practices that include estimation and approximation should be broadly integrated into mathematics curricula for students in Grades 1 – 6 (2006).

Beyond its pedagogical affordances, estimation is a valuable skill for daily life as well as aspects of mathematical work (Bright, 1976; Jones & Taylor, 2010; Sowder, 1992). For example, traffic patrol officers estimate the speed of moving vehicles, tailors estimate the amount of cloth needed to make or alter a garment; land surveyors and landscapers estimate the size of land tracts (Adams & Harrell, 2003). Despite its practical importance and presence in national standards documents, research shows that children and adults exhibit poor measurement estimation performance (Hildreth, 1983; Siegel, Goldsmith, & Madson, 1982). Furthermore, estimation remains largely neglected by researchers (Jones, Gardner, Taylor, Forrester, & Andrew, 2012). For example, little is known about the nature of students' estimation errors, whether students exhibit a propensity to over- or underestimate, or whether estimation accuracy can be improved through instruction.

Quantitative estimation involves computational, numerosity, or measurement estimation (Hogan & Brezinski, 2003; Joram, Subrahmanyam, & Gelman, 1998; Sowder, 1992). Computational estimation involves mental calculation, such as predicting the sum of two multidigit numbers before applying an algorithm or invented strategy. Numerosity estimation involves discriminating, without counting, the number of discrete objects in a group. Measurement estimation involves estimates of continuous extent, such as estimating the square footage of a house. Little research has been conducted specifically on measurement estimation, and most of those studies have focused on length (Joram, Gabriele, Bertheau, Gelman, & Subrahmanyam, 2005; Jones, Gardner, Taylor, Forrester, & Andrew, 2012; Sowder, 1992) or include a variety of physical attributes (Corle, 1960; Gooya, Khosroshahi, & Teppo, 2011) without adequate analysis of specific attributes. Specifically, area measurement estimation performance has been largely ignored by research. For example, when people complete a task such as estimating the square footage of a house, do they think about one unit and how it covers or do they think about an array of square units, or do they think about estimating lengths and multiplying them?

The teaching and learning of many elementary and middle school curricular topics involve area models: the distributive property, multi-digit multiplication, representations of fractions and operations on fractions, as well as decimals. However, children's difficulties in learning area measurement concepts are well documented (Battista, 2004; Kordaki & Potari, 2002; Miller, 2013; Outhred & Mitchelmore, 2000; Piaget, Inhelder, & Szeminksa, 1960). For example, children often use the rectangular area formula without understanding key area measurement concepts such as identifying a unit, coordinating linear and area units, tessellating the plane, and connecting multiplication-as-repeated addition to the notion of an iterated row or column of identical square units to form an array (Battista, 2004; Kara, Eames, Miller, Cullen, & Barrett, 2011; Lehrer, 1998; Outhred & Mitchelmore, 2000). In the US, Grade 8 children's performance in measurement is the weakest of all content areas included on the National Assessment of Educational Progress [NAEP] and the Third International Mathematics and Science Study [TIMSS] (Thompson & Preston, 2004). A typical curricular approach to teaching area measurement procedures in the US does not include the practice of estimating area measurements (Coburn & Shulte, 1986; Smith, Males, & Gonulates, 2016). Taken together, these results suggest that the pedagogical potential of area measurement estimation as a sense-making tool is not being fully realized. Thus, attention to children's area measurement estimation capabilities is long overdue.

Cognitive Processes that Influence Area Estimation Performance

Although area measurement and estimation appear to have a shared conceptual foundation (Bright, 1976), some psychologists argue that the underlying cognitive processes for area measurement estimation and physical area measurement differ (Dehaene, 1997; Hogan & Brezinski, 2003; Joram et al., 1998). In contrast to approaching measurement estimation performance as a skill that is analogous to measurement (Bright, 1976), researchers in psychology have sought to explain the functioning of a rapidly operating, nonverbal, and autonomous system that underlies area estimation performance (see Brannon, Lutz, & Cordes, 2006; Meck & Church, 1983).

A sequence of studies suggests that infants and young children are sensitive to number and area due to systems that quickly and autonomously produce a sense of numerosity or area (see Brannon, Lutz, & Cordes, 2006; Odic, Libertus, Feigenson, & Halberda, 2013; Odic, Pietroski, Hunter, Lidz, & Halberta, 2013). Studies have shown that these approximate number and area representations rely on similar but distinct systems, which are described by Weber's law (Odic, Libertus, et al., 2013; Odic, Pietroski, et al., 2013). Weber's law states that the discriminability of two quantities is not dependent upon their magnitudes or the absolute difference between them, but rather is related to their ratio[1]. For example, both children and adults are fast and accurate at determining whether an array containing 20 blue dots and 10 yellow dots (a ratio of 2.0) contains more blue or yellow dots. However, when shown an image containing 18 blue dots and 15 yellow dots (a ratio of 1.2), participants are slower and make more errors.

Other studies suggest that the cognitive processes that underlie estimation performance involve a nonverbal counting procedure that generates mental magnitudes, which represent quantities. This view predicts that each enumerated area unit is represented by a fixed increment of magnitude added to the contents of a mental accumulator (Cordes, Gelman, Gallistel, & Whalen, 2001; Meck & Church, 1983). There is error in these magnitude representations in memory; that is, they are noisy. The accumulated error, or noise, is proportional to the magnitude. The accumulator model explains the Weber-law characteristic of quantity judgments—namely, the accuracy and speed with which two quantities may be discriminated is

---

[1] We invite the reader to determine a personal Weber fraction for numerosity at www.panamath.org/testyourself.php

determined by their ratio (see Brannon et al., 2006; Cordes, Gelman, Gallistel, & Whalen, 2001; Newcombe et al., 2015; Odic, Libertus, et al., 2013; Odic, Pietroski, et al., 2013).

Findings Related to Children's Area Estimation Capabilities

Results from research suggest that numerosity and measurement estimation form one unique estimation skill (Hogan & Brezinksi, 2003; Joram, Subrahamanyam, & Gelman, 1998; Whalen et al., 1999). Thus, research on numerosity estimation could inform our understanding about area estimation performance. In an early study conducted by Kaufman, Lord, Reese, and Volkman (1949), when undergraduates were shown arrays of more than 10 dots, they tended to underestimate the number of dots shown. Because numerosity and measurement estimation form one unique estimation skill (Hogan & Brezinski, 2003), and because discrete two-dimensional arrays of dots and continuous two-dimensional regions of area are similar in nature, it seems reasonable to hypothesize that area estimation performance should reflect underestimates for regions larger than 10 square units. However, the body of research is silent about this phenomenon as it applies to children and area measurement estimation.

Some studies have shown that performance can be improved for length measurement estimation (Bright, 1979) and area measurement estimation (Hildreth, 1983). Hildreth (1983) examined processes that were associated with success in linear and area estimation for students in Grades 5 and 8, as well as college freshmen. These included mental unit iteration, subdivision, using prior knowledge about the object being measured, comparing the estimated object (or a part of the subdivided object to be estimated) with another object (i.e., using a familiar benchmark), estimating a little bit low and a little bit high and narrowing in on the measure, estimating perpendicular adjacent lengths and multiplying them to estimate area, and rearranging part of the region to estimate its area. Informed by these successful strategies, Hildreth implemented a strategy-based instructional intervention. The students receiving the intervention exhibited more and better strategy use than a comparison group. However, because the researchers focused predominantly on strategy, it is not known whether the strategy-based instruction group exhibited quantitatively improved estimation accuracy.

Relating Area Measurement and Area Estimation

Bright (1976) defined *measurement* as a process of comparing a specified unit and an attribute of an object and *estimation* as a process of arriving at such a comparison without using tools. Defined in this way, area estimation is a process that is analogous to aspects of physical area measurement (Joram et al., 1998) conducted mentally. These processes may include subdivision, which involves mentally partitioning to establish the area unit, or the iteration of area units (or groups of units), which involves transposing an area unit (or group of units) through two-dimensional space to occupy successive positions, always adjacent with one concurrent edge (Miller et al., in press). Furthermore, area measurement is at least partly constructed as a coordination of number operations and spatial operations (citation); thus, area estimation plays a role in area measurement. Area estimation involves applying increasing numerical values in area and mapping from number to area and area to number (c.f., Jones, Forrester, Gardner, Andre, & Taylor, 2012; Joram, Subrahamanyam, & Gelman, 1998).

A learning trajectory approach. Researchers in mathematics education have studied the development of cognitive processes for physical area measurement using a hypothetical learning trajectory (LT) approach (Authors, Date; Simon & Tzur, 2004). An LT is defined in terms of three constituent parts: (a) a domain-specific learning goal; (b) a developmental progression of levels of thinking; and (c) the “instructional tasks designed to engender those mental processes or actions hypothesized to move children through a developmental progression of levels of thinking" (Clements & Sarama, 2004, p. 83). Our recent work (Authors, Date) provided empirical validation of the developmental progression component of an LT for area measurement for children in pre-K through Grade 5 (see Table 1).

-- insert Table 1 --

The first three levels span the early years as young children recognize area as an attribute at the Area Quantity Recognizer level (1), then begin to completely cover a rectangle with physical tiles at the Physical Coverer and Counter level (2), and next develop the ability to draw a complete covering without gaps or overlaps and in approximation of rows at the Concrete Coverer and Counter level (3). The Area Unit Relater and Repeater level (4) represents an important shift in the LT for physical area measurement because this is the level at which key concepts related to unit and unit iteration develop. Levels 5 and 6 represent key mathematical goals for the elementary grades (NGA & CCSSO, 2010). The concept of unit iteration is extended as children next learn to identify a square unit as a component of a unit of units (i.e., a row or column of identical units) at the Initial Composite Structurer (5) level. Eventually, children see a square unit as a component of a unit of unit of units (i.e., as an element of an array composed of rows and columns of identical square units) at the Area Row and Column Structurer (6) level. Levels 7 and 8 are relevant as children transition into the upper elementary grades and middle school when area formulas for other shapes are developed.

Because area measurement is a key mathematical goal in the elementary grades, we targeted our inquiry on levels of the LT for area measurement that are most relevant to the elementary grades: the ICS and ARCS levels. Furthermore, because we sought to link the findings of the present study related to area estimation to an LT for area measurement, we designed an instructional intervention that we hypothesized could support students in refining their area estimates using area estimation strategies demonstrated by successful estimators (Hildreth, 1983) and area measurement concepts that are familiar to children who are operating predominantly at the ICS (recognizing a unit of units) and ARCS (recognizing a unit of unit of units) levels. Thus, we conjectured that feedback on the numerical value of the area of a rectangular region, in tandem with a visual image showing a rectangular region being tiled by a row followed by groups of rows of unit squares, would support students' area estimation performance. We address the following research questions:

1. How well do children at the Initial Composite Structurer and Area Row and Column Structurer levels of the LT for area measurement estimate areas of rectangles and what is the nature of their estimates?
2. For children operating predominantly at the ICS and ARCS level, to what extent can their area estimation be quantitatively improved?

Method

Participants

We recruited a sample of 42 children from three different Grade 4 classrooms at a suburban public school in the Midwest. The classes of size 21, 22, and 23 students were taught by different teachers, and the Everyday Math series by McGraw Hill was used in each class. All children who returned a parental consent letter to the teacher were included in the study.

We targeted Grade 4 because our prior work indicated that most Grade 4 children in a typical US educational context operate at the ICS and ARCS levels of the LT for area measurement (Authors, Date). The 42 children were randomly assigned to create two equal-size groups, a treatment and a control group, each of size 21.

The children in the treatment group engaged in a total of five one-on-one area estimation sessions with a researcher: a pretest and three instructional sessions conducted over 4 consecutive school days, and a posttest conducted approximately 1 month after the third instructional session. Children in the comparison group took part in only two area estimation sessions with a researcher, a pretest session and a posttest session conducted approximately one month later. A conceptual area measurement written assessment was administered before the first and after the final area estimation feedback sessions. We were unable to collect a complete data set for four of the initial sample of 42 children due to absence from school. Therefore, our final sample of 38, which was subjected to complete analysis, included a treatment group of 20 children and a comparison group of 18 children.

Instrumentation

Data sources include children's verbal numerical responses from the area estimation sessions as well as their responses to the conceptual area measurement written assessment both times that it was administered.

Conceptual area measurement written assessment. The conceptual area measurement written assessment was designed to provide evidence of each child's predominant area LT level to increase the potential fit between their level of understanding and the instructional support. Thus, the assessment consisted of three items that differentiated the levels of an LT for area measurement (Authors, Date). Those concepts and processes include iterating units of area, constructing and operating on composite units of area, and coordinating linear and area units to compose composite units (units of units) of area.

Area estimation pretest session. The area estimation pretest session consisted of a total of 24 trials. Each trial was comprised of a single estimation task presented on a computer screen.

The child was shown a nonstandard unit square, labeled 1 square unit[3], with a rectangle on the same screen. A researcher asked, “If this small square has an area of one, what is the area of this rectangle?” The researcher typed the child’s response into an answer box on the computer without feedback. Each trial paired the same unit square with rectangles ranging from four to 200 square units. To control for a possible effect of rectangle size on area estimation performance, we utilized a total of four small (20 or fewer square units), 12 medium, and eight large rectangles (more than 80 square units).

Targeted Feedback: Sessions 1, 2, and 3

The three instructional sessions each consisted of a total of 28 trials, which involved repeating the same procedure and question as the estimation pretest session; however, the first 16 of these trials also involved targeted feedback. This targeted feedback was designed to include area estimation strategies demonstrated by successful estimators (Hildreth, 1983) and area measurement concepts that are familiar to children who are operating predominantly at the ICS and ARCS levels of the area LT, as well as support for encouraging the child to reflect on his or her estimate. Specifically, the feedback consisted of three parts: (a) a dynamic display of the iteration of a single square unit across the rectangle to create a row, which was then iterated up to completely cover the rectangle with rows and columns of identical square units, (b) the actual area of the rectangle in numerical format; and (c) the percent error between the actual area of the rectangle and the child’s estimate. As the first row was completed, a report of the total number of square units that fit in the row was displayed outside of the rectangle just to the right of the row. With each iteration of the row, the numeric display updated to state the number of square units covering the rectangle “so far.” Each of the three instructional sessions included 16 trials with feedback, followed by 12 trials with no feedback provided (which were used to identify changes in children's estimation performance).

Area estimation posttest session. The area estimation posttest session was identical in structure to the area estimation pretest session. The posttest was administered to both treatment and comparison groups, allowing us to isolate the effects of the instructional sessions on the treatment group.

Analysis

Evaluating conceptual area measurement understanding. Two members of the research team collaboratively coded children's responses to the conceptual area measurement written assessment using an LT for area measurement (Authors, Date). Based on the observable behaviors evident in children's written responses, the two researchers reached consensus for the mapping of responses for each of the three items to a level of the LT for area measurement.

The conceptual area measurement pretest revealed that the treatment group (*n* = 20) consisted of two children who were operating predominantly at the ARCS level, 17 children at the ICS level, and 1 child for whom we could make no predominant level claim[4] based on the observable behaviors in the child's written responses. The comparison group (*n* = 18) consisted of four children operating predominantly at the ARCS level, 12 children at the ICS level, and 2 children for whom no level claim could be made.

Evaluating area estimation performance. Analysis of the area estimation sessions proceeded according to two phases involving (a) describing children's area estimation performance, and (b) looking for changes in children's area estimation performance. Describing children's area estimation performance presents inherent methodological difficulties (Sowder, 1992) related to various methods for evaluating area estimation performance. One method involving determining ranges for “reasonable” or “incorrect” responses for estimation performance requires the researcher to judge what might seem like a good estimate. This approach presents a challenge for synthesizing results across a multitude of studies when research teams define different performance categories.

Percent error provides a potential solution to this difficulty in *describing* area estimation performance; however, this approach presents a new set of challenges when used to *compare* area estimation performance across items or participants. If percent error for underestimates is defined as negative and percent error for overestimates as positive, then combining percent error scores to find an average would allow for the loss of overestimates and underestimates by a zero-sum process. This could be avoided by calculating separate averages in percent error for under- and overestimates for a particular participant or item. Alternatively, absolute percent error has the potential to alleviate the potential difficulty of cancellation of under- with overestimates, but percent error for underestimates is naturally bounded between -100% and 0% whereas percent error for overestimates is unbounded. Furthermore, it appears intuitively unfair that an underestimate of 90% should be considered as the same level of performance as an overestimate of 90%. For example, an estimate of 1 square unit for a 10-square unit rectangle seems like a poorer area estimation performance than an estimate of 19 square units for a 10 square unit rectangle. The underestimate was one tenth of the actual area of the rectangle, but the overestimate was just short of double the actual area.

In response to these methodological difficulties, we derived separate functions for transforming participants’ under- and overestimates to area estimation “penalty scores.” We calculate the penalty score for overestimates using the formula $O(e, a) = 5(e/a - 1)$, where e is estimate and a is the actual area of the rectangle. Scaling by a factor of 5 allows for each unit increase in penalty score to correspond to 20% error. We chose this scale because we conjectured that participants' estimates would not be much better than 20%. Similarly, we used the formula $U(e, a) = 5(a/e - 1)$ for underestimates. This scaling allows for an underestimate penalty score of 1 to correspond to a percent error of approximately 17%. With this transformation, penalty scores are related to the ratio between the actual and estimated areas, all penalty scores are positive, and penalty scores associated with both under- and overestimates are generated by unbounded functions.

Results and Discussion

We present our findings with respect to our two research questions: 1) How well do children at the ICS and ARCS levels of the LT for area measurement estimate areas of rectangles and what is the nature of their estimates? And 2) To what extent can ICS and ARCS level children's area estimation performance be quantitatively improved through targeted instruction emphasizing spatial structuring with numeric feedback? In the following sections we first describe fourth-grade children's area estimates of rectangles including the direction of errors. We found a clear tendency to make underestimates rather than overestimates. Next, we describe the changes in children's estimates of areas of rectangles. Children showed marked improvement after as little as two sessions.

The Nature of Children's Estimation of Rectangular Area

Tendency to underestimate. We analyzed all 38 children's verbal numerical responses to the initial 24-item area estimation pretest. We first calculated percent error for each response and identified each as an under- or overestimate. Next, we constructed Figure 1 to depict the nature of the distribution of children's estimates of rectangle areas.

-- insert Figure 1 --

We found a right-skewed distribution with far more instances of underestimates (671) than overestimates (241) with this sample of Grade 4 children. A two-tailed, one-sample *z*-test revealed that children provided significantly more underestimates than a null outcome of 50% with $z = 14.24$ and $p < .0001$. This suggests that children are more inclined to underestimate area measures of rectangles, given the rectangle sizes and shapes used in this study.

Identifying outliers. We considered individual responses as outliers and removed them if they had a substantial effect on the descriptive statistics. Inspection of the data across all sessions revealed three extreme points, each from a different child. For example, on the pretest, for a rectangle with an area of approximately 111 square units, a child in the treatment group gave an estimate of 5 square units. This corresponds to a percent error of -95% and a penalty score of 106. We removed these three data points.

Individual differences in area estimation performance. After noticing the prevalence of underestimates across the entire sample of 38 children we looked for differences in estimation tendencies by looking at each individual's 24 estimates for the area estimation pretest. We created categories, finding only two children purely overestimated (0 to 3 underestimates out of 24 possible), no children primarily overestimated (4 to 7 underestimates), nine children gave a mix of over and underestimates (8 to 15 underestimates), 10 primarily underestimated (16 to 19 underestimates), and the majority, 17 purely underestimated (20 to 24 underestimates). We note that the two children who purely overestimated were in contrast with the group-level findings.

Improving Children's Area Estimation Performance

We examined changes in children's estimation performance throughout the study by comparing 95% confidence intervals surrounding the mean penalty scores across the pre- and posttests for each group, as well as the 12-item assessments that followed the feedback sessions (Sessions 1, 2, and 3) for the treatment group (Figure 2a and 2b).

-- insert Figures 2a and 2b --

Figure 2a illustrates the rate of growth of the treatment group across the sessions. From the Pretest to Session 1, the mean penalty score decreased; however, the 95% confidence intervals for the mean penalty score overlap. By Session 2 this confidence interval no longer overlaps with the pretest confidence interval. This suggests that children in the treatment group demonstrated significantly improved area estimation performance by Session 2 (over their performance on the pretest), and maintained that improved performance throughout the study. In addition, the confidence intervals are markedly shorter for the treatment group in Sessions 2, 3 and the Posttest than those for the Pretest and Session 1. Taken together, these findings suggest that, after treatment, children's estimation errors were less variable, significantly reduced, and remained at that level through the Posttest.

Figure 2b shows 95% confidence intervals for mean penalty score for the treatment and comparison groups for the Pretest. The comparison and treatment groups exhibit overlapping 95% confidence intervals for mean penalty score that are approximately the same width. This suggests that the estimation performance of the two groups is not significantly different at the outset of the study. Furthermore, there is no overlap between the 95% confidence intervals for mean penalty score for the treatment and comparison groups in the Posttest session. In addition, the width of the 95% confidence interval for the treatment group is remarkably narrower than the 95% confidence interval for the comparison group. This provides further evidence that the treatment had a significant effect on children's area estimation performance.

Effect of treatment relative to rectangle size. To investigate the effect of the feedback on children's area estimation performance for rectangles of different sizes, we compared 95% confidence intervals surrounding the mean within each rectangle size (Figure 3) category across the Pre- and Posttests for the treatment group.

-- insert Figure 3 –

Figure 3 shows that pre- and posttest 95% confidence intervals for the mean penalty scores for medium and large rectangles do not overlap. This suggests that children in the treatment group exhibited significant improvement in area estimation performance within the medium and large rectangle size categories. Furthermore, the width of the confidence intervals generally decreased from pre- to posttest for the medium and large rectangle size categories, which indicates that the variability in students' estimates was reduced. However, there is no overlap in the penalty scores for the small and medium as well as the small and large rectangle categories during the posttest. This indicates that, although children's area estimation performance was improved for the medium and large rectangle categories, this improvement did not reach the level at which they could estimate areas of small rectangles (which remained unchanged throughout the study).

Effect of treatment on propensity to underestimate. Our analysis of the pretest data revealed children's propensity to underestimate areas of rectangular regions (see Figure 1). Specifically, the treatment group underestimated areas of rectangles 357 out of 480 instances on the pre-test. After the feedback sessions, the treatment group underestimated 342 out of 480 instances on the posttest. A two-tailed two-proportion *z*-test indicates that this difference in frequency of underestimates is not significant. This suggests that, although the children in the treatment group exhibited improved area estimation performance along multiple dimensions as described above, the propensity to underestimate persists even with instruction.

Conclusions and Implications

We investigated the nature of ICS and ARCS level children's performance for estimating areas of rectangular regions, as well as the effects of an instructional intervention on their estimation performance, which was based on building and repeating rows of identical square units and giving numerical feedback. Due to small size of the sample, we regard the conclusions based on our findings presented in the sections below as researchable conjectures.

The Nature of Children's Performance in Estimating Areas of Rectangles

The findings of the present study provided empirical validation for a hypothesis supported by results from prior research on estimation performance in mathematics education and psychology. If area measurement estimation is the same skill as numerosity estimation, as suggested by Hogan and Brezinski (2003), and if adults tend to underestimate numerosity for arrays of 10 to 210 dots (Kaufman, Lord, Reese, & Volkman, 1949), then area estimation performance should reflect underestimates for regions larger than 10 square units. For the area estimation pretest, we observed the ratio of Grade 4 children's underestimates to overestimates to be approximately 3:1. For the treatment group, the ratio of underestimates to overestimates was not significantly different even after the group demonstrated significantly improved area estimation performance in response to the treatment. These results provide strong evidence in favor of this hypothesis – that children would underestimate areas of rectangular regions – as it applies to ICS and ARCS-level children and area estimation performance.

Although the sample as a whole exhibited a general tendency to underestimate areas of rectangular regions, we found some individual students exhibited estimation performance that differed from the group trend. That is, two children in the sample exhibited a tendency to overestimate, whereas 17 children in the sample exhibited a tendency to underestimate. This suggests our finding that children exhibited an overall tendency to underestimate, as exhibited by Figure 1, may describe most children, but does not accurately describe every child. We wonder if the purely overestimating children are simply not as cautious in disposition as the underestimating children. More research is needed to explore possible reasons why a small fraction of children exhibit trends in estimation performance that differs from their peers.

Improving Children's Performance for Estimating Areas of Rectangles

We designed the intervention to improve ICS and ARCS-level children's area estimation performance utilizing area measurement strategies that were familiar to them. Specifically, we designed the intervention to support children in engaging in the deliberate and logical mental action of *building* a row of identical square units and *iterating* that row to form an array of identical rows and columns of square units. These are mental actions and objects that we associate with the ICS and ARCS levels of the area LT. This study employed this visual structural support accompanied by numerical feedback. We observed a significant effect on children's performance for estimating areas of rectangular regions. However, we cannot isolate the effects of the visual structural support from the effects of numerical feedback (in the form of reports of the number of square units in a row, and the number of square units covering the rectangle "so far" as the row was iterated to cover the rectangle) or feedback about the efficacy of their estimates (in the form of percent error associated with the child's estimate). Future research should explore the effectiveness of numerical feedback with a separate consideration for visual structural support.

Children's performance for estimating areas of small rectangles, with a mean penalty score of 1.58, corresponding to a percent error of approximately 32% for overestimates and approximately 24% for underestimates (see Figure 3), may indicate optimum performance for estimating areas of rectangles. These might lay the foundations for selecting benchmarks for what constitutes a response that is "reasonable" or close enough to the actual area of a rectangle to be considered correct. The percent error scores reported here could alternatively be interpreted as a ratio of children's internally represented area, demonstrated by an estimate, to the actual area of the rectangle. A 32% difference in the internally represented magnitude to the actual magnitude would then correspond to a nearest whole number ratio of 3:2. Similarly, a 24% difference between internally represented magnitude and actual magnitude for underestimates would correspond to a nearest whole number ratio of 4:5. These ratios are similar in magnitude to empirically estimated Weber fractions that govern the discriminability of two regions of area for young children (Odic, Libertus, et al., 2013). Future research is needed to explore if it is possible for children or adults to become more accurate than 32% over and 24% under when estimating areas of rectangular regions.

In the present study, we looked for rate of growth between the estimation pretest session and Sessions 1, 2, 3, and the estimation posttest session. When would the change in rectangular area estimation performance first be significant? The only pairwise significant differences involved comparisons between the pretest and Sessions 2 and 3, and the posttest (see Figure 2a). Therefore, the length of the treatment, a total of three sessions, was enough to bring the children's level of error in area estimation to a plateau. This suggests that improving area estimation performance takes at least two instructional sessions, with at least 24 trials per session.

The children in the treatment group exhibited significantly improved area estimation performance for the medium and large rectangle size categories, with the widest margin of improvement in the large rectangle size category. However, even on the area estimation posttest, children demonstrated a significantly better performance when estimating areas of small rectangles than medium and large rectangles. These results suggest that the capacity to improve area estimation performance is dependent upon the size of the ratio between the estimated object and the unit used to measure. We note that in some real world and “school math” settings the ratio between the unit of measure and the object to be measured is large, but a large standard unit of area is not necessarily conventionally employed. For example, square feet or square meters are typically used to measure floor space in a home, and acres are used to measure two dimensional space in a field or on a farm. It would not be extraordinary to observe measures of area that exceed 100 or 1000 in these contexts. Furthermore, it would not be unusual for the answer to a typical “school math” area measurement problem to have values in our medium and large rectangle categories. Thus, the fact that area measurement estimation performance declines as we attempt to estimate relatively large areas has both practical and pedagogical significance.

Findings suggest that the treatment significantly reduced the occurrences of extreme responses. We hypothesize that the feedback helped children in the treatment group to develop benchmarks for the smallest and largest rectangles appearing on the computer screen, and that these benchmarks served as an effective cognitive tool for children in the treatment group to support them in adjusting their estimates of areas of rectangular regions (Bright, 1976; Gooya et al., 2011). This finding suggests that the intervention employed here, which made use of familiar area measurement strategies, engaged children’s rapidly operating and autonomous representation system as well as their deliberate logical sense-making system. Moreover, we suspect that engaging children’s deliberate and logical systems was necessary for improving area estimation performance.

Area Estimation Performance and Sophistication in Conceptual Area Measurement

We designed the treatment employed here to support children's use of a mental image of a composite quantity for area to build structured regions from composites of known quantities; these are concepts and processes that are familiar to children operating at least at the ICS level of the LT for area measurement. We speculate that the children, all[7] of whom were operating at least at the ICS level, readily responded to the intervention because they possessed the necessary cognitive components for enacting a mental structuring operation using benchmark quantities (i.e., units and composite units). Future research should more fully explore the pedagogical potential for including estimation as a sense-making tool (ACARA, 2014; Department for Education, 2013; NCTM 1989, 2000; NGA & CCSSO, 2010) for the teaching and learning of area measurement. Specifically, future studies are needed to investigate whether younger children who are operating predominantly at levels of the area LT below the ICS level, could develop significantly improved area estimation performance *and* conceptual area measurement understanding from such an intervention.

Acknowledgements

We would like to thank Pamela Beck her valuable contributions to the design of this study and data collection.

Footnotes

[1] We invite the reader to determine a personal Weber fraction for numerosity at www.panamath.org/testyourself.php

[2] This unit varied, depending on the size of the computer screen used and the window magnification, but the unit square was close to 1 square centimeter in area (+- 90%). Given the U.S. student population sampled, we believe the unit was non-standard in that it was not a square inch.

[3] No level claim was made when a child's observable behaviors were not indicative of a particular level of the LT for area measurement.

[4] The median percent error for all underestimates was approximately 38%, and the median percent error for all overestimates was approximately 25%.

[5] Because the distributions of over- and underestimates were skewed (see Figure 1), we employed the median as a group measure of the children's average area estimate for each rectangle in the regression analyses. However, when comparing area estimation performance across groups, we turn to comparing interval estimates surrounding the mean.

[7] The set of responses of three children, two in the comparison group and one in the treatment group, for the written conceptual area measurement assessment did not yield observable behaviors that the research team could map to the LT for area measurement. Thus, no predominant level claim was made for three of the 38 participants.

References

Adams, T. L., & Harrell, G. (2003). Estimation at work. In D. H. Clements & G. Bright (Eds.), *Learning and teaching measurement: 2003 yearbook* (pp. 229–244). Reston, VA: National Council of Teachers of Mathematics.

Australia Curriculum, Assessment and Reporting Authority [ACARA], 2014. *Foundation to year 10 curriculum: Mathematics*. Retrieved from http://www.australiancurriculum.edu.au/mathematics/curriculum/f-10?layout=1

Barrett, J. E., Clements, D. H., Sarama, J., Miller, A. L., Cullen, C. J., Van Dine, D. W., …Klanderman, D. (2017). Integration of results: A revised learning trajectory for area measurement. In J. E. Barrett, D. H. Clements, & J. Sarama (Eds.), *Children's measurement: A longitudinal study of children's knowledge and learning of length, area, and volume* (JRME Monograph No. X, pp. 129–148). Reston, VA: National Council of Teachers of Mathematics.

Brannon, E. M., Lutz, D., & Cordes, S. (2006). The development of area discrimination and its implications for number representation in infancy. *Developmental Science*, *9*(6), F59–F64. doi:10.1111/j.1467-7687.2006.00530.x

Battista, M. T. (2004). Applying cognition-based assessment to elementary school students' development of understanding of area and volume measurement. *Mathematical Thinking and Learning*, *6*(2), 185–204. doi:10.1207/s15327833mtl0602_6

Bright, G. W. (1976). Estimation as part of learning to measure. In D. Nelson & R. E. Reys (Eds.), *Measurement in school mathematics: 1976 yearbook* (pp. 87–104). Reston, VA: National Council of Teachers of Mathematics.

Bright, G. W. (1979). Measuring experienced teachers' linear estimation skills at two levels of abstraction. *School Science and Mathematics*, *79*, 161–164. doi:10.1111/j.1949-8594.1979.tb09469.x

Cheung Kwok-cheung. (2004). *Full-time Obligatory Education: Mathematics Curriclum Standards (China)*. Retrieved from http://ncm.gu.se/media/kursplaner/andralander/kinagrund.pdf:

Clements, D. H., & Sarama, J. (2004). Learning trajectories in mathematics education.

Mathematical Thinking and Learning, *6*(2), 81–89. doi:10.1207/s15327833mtl0602_1

Clements, D. H., Barrett, J. E., Sarama, J., Cullen, C. J., Van Dine, D. W., Eames, C. L., …Vukovich, M. (2017). Length – A summary report. In J. E. Barrett, D. H. Clements, & J. Sarama (Eds.), *Children's measurement: A longitudinal study of children's knowledge and learning of length, area, and volume* (JRME Monograph No. X, pp. 25–68). Reston, VA: National Council of Teachers of Mathematics.

Cordes, S., Gelman, R., Gallistel, C. R., & Whalen, J. (2001). Variability signatures distinguish verbal from nonverbal counting for both large and small numbers. *Psychonomic Bullentin & Review*, *8*(4), 698–707. doi:10.3758/BF03196206

Corle, C. G. (1960). A study of the quantitative values of fifth and sixth grade pupils. *The Arithmetic Teacher*, *7*(7), 333–340.

Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York, NY: Oxford.

Department for Education (2013). *The National Curriculum in England: Key Stages 1 and 2 Framework Document*. Retrieved from https://www.gov.uk/government/publications/national-curriculum-in-england-mathematics-programmes-of-study

Gooya, Z., Khosroshahi, L. G., & Teppo, A. R. (2011). Iranian students' measurement estimation performance involving linear and area attributes of real-world objects. *ZDM*, 43, 709–722. doi:10.1007/s11858-011-0338-1

Hildreth, D. J. (1980). *Estimation strategy uses in length and area measurement tasks by fifth and seventh grade students* (Doctoral dissertation). Available from ProQuest Dissertations & Theses database. (Order No. 8107340).

Hildreth, D. J. (1983). The use of strategies in estimating measurements. *Arithmetic Teacher*, *30*(5), 50–54.

Hogan, T. P., & Brezinski, K. L. (2003). Quantitative estimation: One, two, or three abilities? *Mathematical Thinking and Learning*, *5*(4), 259–280. doi:10.1207/S15327833MTL0504_02

Jones, M. G., & Talyor, A. (2010). Developing a sense of scale: Looking backward. *Journal of Research in Science Teaching*, *46*(4), 460–475.

Jones, M. G., Gardner, G. E., Taylor, A. R., Forrester, J. H., & Andrew, T. (2012). Students' accuracy of measurement estimation: Context, units, and logical thinking. *School Science and Mathematics*, *112*(3), 171–178. doi:10.1111/j.1949-8594.2011.00130.x

Joram, E., Gabriele, A. J., Bertheau, M., Gelman, R., & Subrahmanyam, K. (2005). Children's use of the reference point strategy for measurement estimation. *Journal for Research in Mathematics Education*, *36*(1), 4–23.

Joram, E., Subrahmanyam, K., & Gelman, R. (1998). Measurement estimation: Learning to map the route from number to quantity and back. *Review of Educational Research*, *68*(4), 413–449. doi:10.3102/00346543068004413

Kara, M., Eames, C. L., Miller, A. L., Cullen, C. J., & Barrett, J. E. (2011). Developing an understanding of area measurement concepts with triangular units. In L. R. Wiest, & T. Lumberg (Eds.), *Proceedings of the 33rd Annual Conference of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 1015 – 1023). Reno, NV: University of Nevada, Reno.

Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkmann, J. (1949). The discrimination of visual number. *American Journal of Psychology*, *62*(4), 498–525. doi:10.2307/1418556

Kordaki, M., & Potari, D. (2002). The effect of area measurement tools on student strategies: The role of a computer microworld. *International Journal of Computers for Mathematical Learning*, *7*(1), 65–100. doi:10.1023/A:1016051411284

Lehrer, R., Jenkins, M., & Osana, H. (1998). Longtudinal study of children's reasoning about space and geometry. In R. Lehrer & D. Chazan (Eds.), *Designing learning environments for developing understanding of geometry and space* (pp. 137–167), Mahwah, NJ: Lawrence Erlbaum.

Meck, W. H., & Church, R. M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes*, *9*(3), 320–324. doi:10.1037/0097-7403.9.3.320

Members of National Focus Group on Teaching of Mathematics India. (2006). *Position Paper for National Focus Group on Teaching of Mathematics (India)*. Retrieved from New Delhi, India: http://www.ncert.nic.in/new_ncert/ncert/rightside/links/pdf/focus_group/math.pdf

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

National Governors Association Center for Best Practices & Council of Chief State School Officers (2010). *Common core state standards for mathematics*. Washington, DC: Author.

Newcombe, N. S., Levine, S. C., & Mix, K. S. (2015). Thinking about quantity: The intertwined development of spatial and numerical cognition. *Cognitive Science*, *6*(6), 491–505. doi:10.1002/wcs.1369

Odic, D., Libertus, M. E., Feigenson, L., & Halberta, J. (2013). Developmental change in the acuity of approximate number and area representations. *Developmental Psychology*, *49*(6), 1103–1112. doi:10.1037/a0029472

Odic, D., Pietroski, P., Huner, T., Lidz, J., & Halberta, J. (2103). Young children's understanding of "more" and discrimination of number and surface area. *Journal of Experimental Psychology*, *39*(2), 451–461.

Outhred, L. N., & Mitchelmore, M. C. (2000). Young children's intuitive understanding of rectangular area measurement. *Journal for Research in Mathematics Education*, *31*(2), 144–167. doi:10.2307/749749

Piaget, J., Inhelder, B., & Szeminska, A. (1960). *The child's conception of geometry*. New York, NY: Basic Books.

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

Siegel, A. W., Goldsmith, L. T., & Madson, C. R. (1982). Skill in estimation problems of extent and numerosity. *Journal for Research in Mathematics Education*, *13*(3), 211–232. doi:10.2307/748557

Simon, M., & Tzur, R. (2004). Explicating the role of mathematical tasks in conceptual learning:
An elaboration of the hypothetical learning trajectory. *Mathematical Thinking and Learning*, *6*(2), 91–104.

Smith, J. P., Males, L. M., & Gonulates, F. (2016). Conceptual limitations in curricular presentations of area measurement: One nation's challenges. *Mathematical Thinking and Learning*, *18*(4), 239–270. doi:10.1080/10986065.2016.1219930.

Sowder, J. (1992). Estimation and number sense. In D. A. Grouws (Ed.), *Handbook of research*
on mathematics teaching and learning (pp. 371–389). New York, NY: Macmillan Publishing Company.

Takahashi, A., Watanabe, T., & Yoshida, M. (2008). *The English Translation of the Japanese*
Mathematics Curricula in the Course of Study. Retrieved from http://ncm.gu.se/media/kursplaner/andralander/Japanese_COS2008Math.pdf

Thompson, T. D., & Preston, R. V., (2004). Measurement in the middle grades: Insights from
NAEP and TIMSS. *Mathematics Teaching in the Middle School*, *9*, 514–519.

Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science*, 10, 130–137.

Table 1

Summary of an LT for area measurement.

Level	Developmental Progression Level Summary
1. Area Quantity Recognizer (AQR)	Perceives two-dimensional space and objects within the space. May employs internal innate competencies to compare extent of regions. Compares regions of area by comparing linear dimensions or summations of the linear dimensions of the regions or objects.
2. Physical Coverer and Counter (PCC)	Visualizes that regions can be covered by other regions when provided with perceptual support. Can direct the covering of space with physical tiles and recognize that covering as complete.
3. Complete Coverer and Counter (CCC)	Applies an explicit understanding that the entire region must be covered with approximately rectangular shapes. Uses a counting-all action scheme to enumerate the covering of a region.
4. Area Unit Relater and Repeater (AURR)	Relates size and number of units (fewer larger, more smaller). Uses rows as an intuitive structure when enumerating the covering of a region. Has developed the concept of unit iteration.
5. Initial Composite Structurer (ICS)	Builds, maintains, and manipulates mental images of composite units, structuring them as composites of individual shapes as a single entity. Applies a composite unit repeatedly, but not necessarily exhaustively
6. Area Row and Column Structurer (ARCS)	Applies a composite unit repeatedly and exhaustively, coordinating movement in one-to-one correspondence with the elements of the orthogonal column. Applies the concept that the length of a segment specifies the number of units that will fit along that segment.
7. Array Structurer (AS)	Has an abstract understanding of the rectangular area formula. Has an internalized mental image of an array of identical square units. Each square can be viewed as a unit, a component of a unit of units (a row or column), and a component of a unit of unit of units (the array).
8. Conceptual Area Measurer	The rectangular area formula is generalizable as area formulas for other shapes are developed

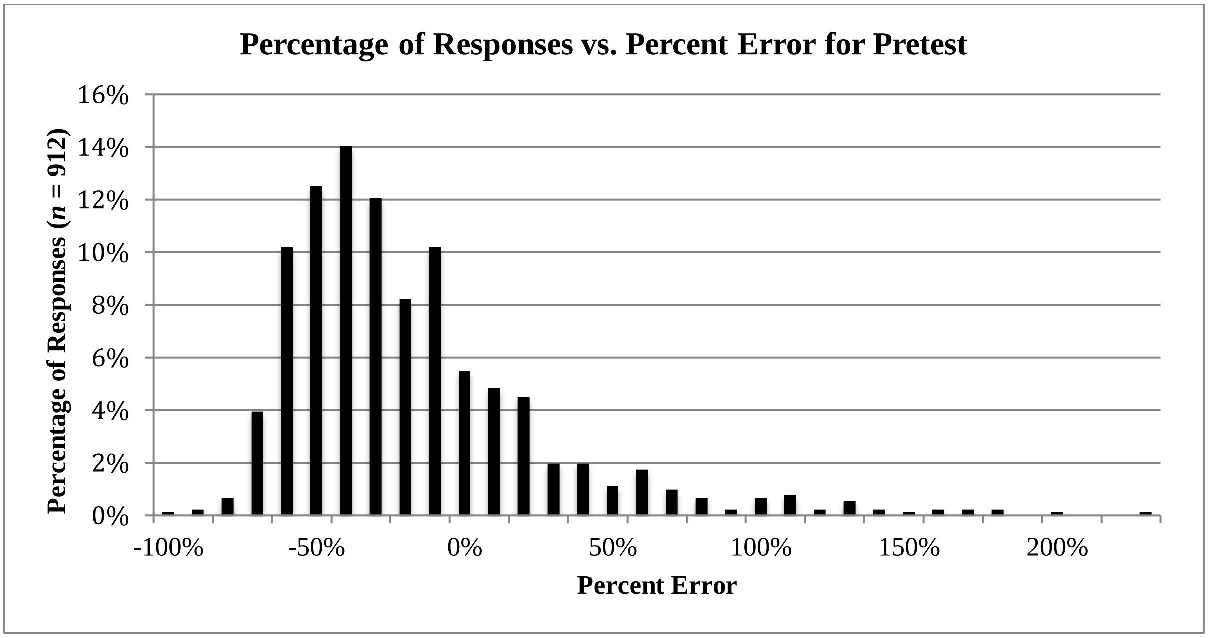

Figure 1. The horizontal axis shows raw percent error and the vertical axis shows the percentage of all responses to the area estimation tasks administered during the pretest session (24 tasks) for all children (n = 38), for a total of 912 responses.

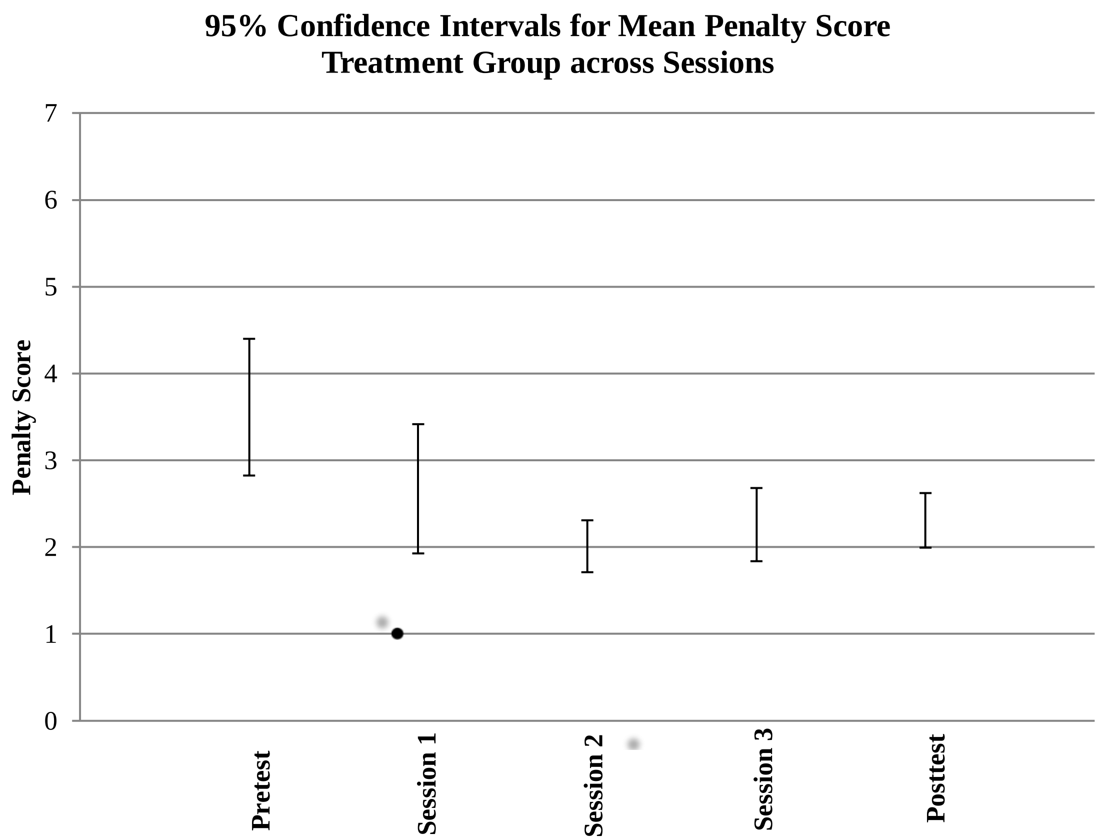

Figure 2a. Treatment Group 95% Confidence Intervals for Mean Penalty Scores.

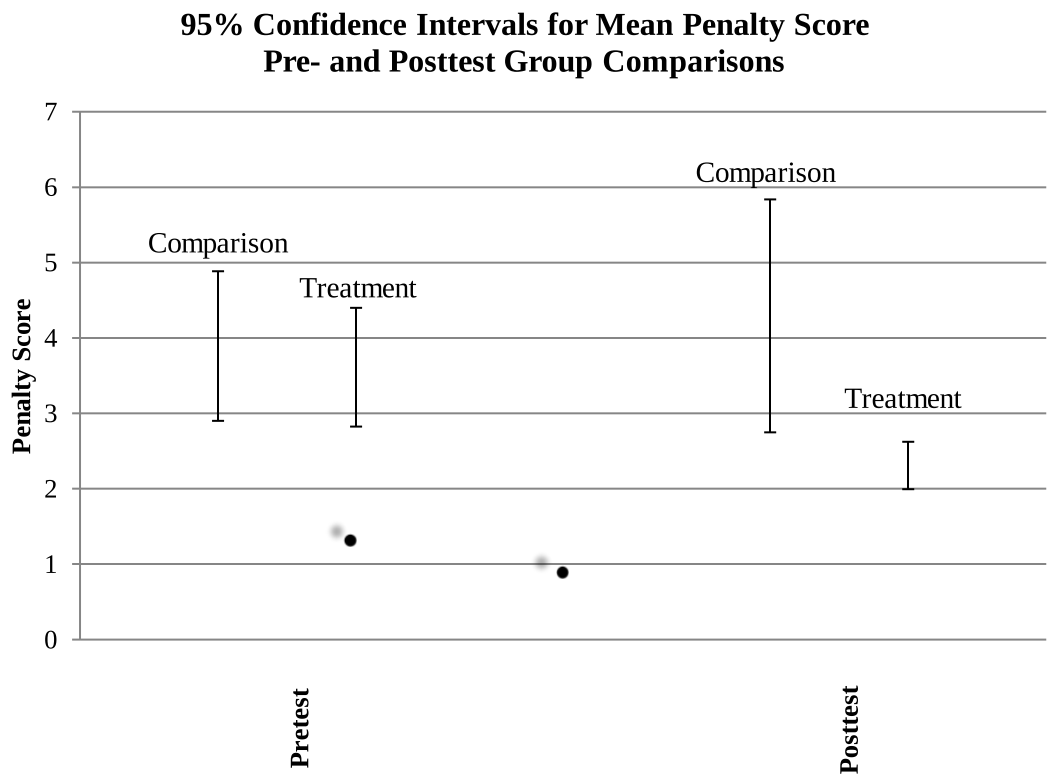

Figure 2b. 95% Confidence Intervals for Mean Penalty Scores: Pre- and Posttest Comparison.

95% Confidence Intervals for Treatment Group Pre- and Posttest Mean Penalty Scores per Rectangle Size Category

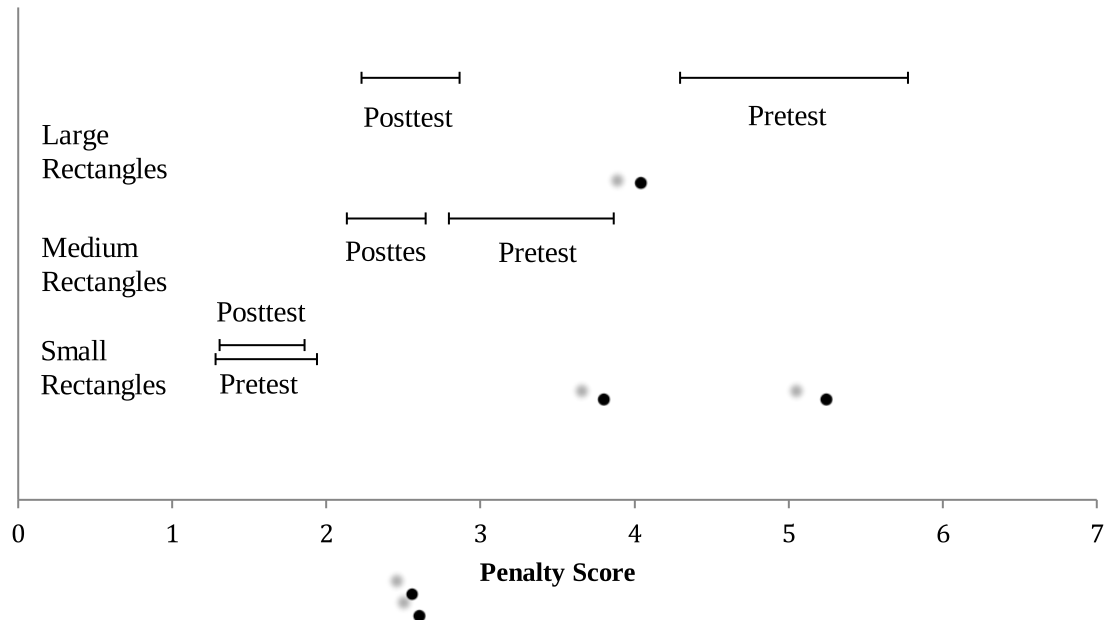

Figure 3. 95% Confidence intervals for pre- and posttest mean penalty scores for small, medium and large rectangle size categories.