Smart Errors in Learning Multidigit Number Meanings

Corinne A. Bower

California State University, Los Angeles

Kelly S. Mix

University of Maryland

Lei Yuan

University of Colorado, Boulder

Gregory R. Hancock

University of Maryland

Linda B. Smith

Indiana University

Abstract

Children's early accuracy on place value (PV) tasks longitudinally predicts their later multidigit calculation skills. However, another window into children's emerging base-ten concepts is the pattern of errors—'smart errors'—they exhibit on these measures. Past research has speculated that these smart errors—similar to invented spelling—might reflect children's initial PV understanding that might be important for later learning of multidigit numbers and calculation. The current study examines the development of smart errors on Base-Ten Counting (invented counting errors) and Transcoding (expanded errors) in 279 U.S. kindergartners (M_{age}=5.76 years) and investigated whether the presence of smart errors is associated with 1) higher concurrent levels of PV task accuracy, 2) greater growth in PV understanding over one year, 3) higher levels of multidigit calculation in second grade. Results indicate that the two errors emerged in an overlapping waves pattern, with expanded errors appearing first and waning earlier than invented counting errors. Kindergartners who made invented counting errors but not expanded errors demonstrated the highest overall concurrent PV understanding. Second, kindergartners who made Transcoding expanded errors showed the greatest growth in PV understanding compared to those who exhibited only invented-counting errors. Third, kindergartners who made invented counting errors alone showed stronger multidigit calculation skills in second grade compared to those who made neither error. Thus, these smart errors reflect partial structural knowledge of place value that is a potentially important developmental contributor to learning multidigit number meanings.

Keywords: *place-value understanding, place-value errors, multidigit numbers, mathematical learning*

Smart Errors in Learning Multidigit Number Meanings

Multidigit numerals represent quantities using base-ten units that are arranged in multiples of ten from right to left, with the digits in each spatial position, or "place," indicating the count of each unit. For example, "253" stands for two sets of 100, five sets of 10, three sets of 1, or [2 × 100] + [5 × 10] + [3 × 1]. Base-ten units and their counts are also represented in verbal number names for these quantities (e.g., "two hundred fifty-three"). Children—specifically first graders—who understand these symbolic representations go on to better mathematics performance throughout elementary and middle school (Chan et al., 2014; Gervasoni et al., 2007; Hiebert & Wearne, 1996; Moeller et al., 2011). However, base-ten symbols—i.e., written digits and corresponding number names—are notoriously difficult for children to grasp (Carpenter et al., 1996; Fuson, 1990; Kamii, 1986), likely because their shared relational structures are not transparent or easily aligned (Mix et al., 2019).

Despite these obstacles, children actively construct multidigit number meanings years earlier than researchers and educators previously thought (*Authors*, 2022b; Byrge et al., 2014; Mix et al., 2014; Yuan et al., 2019). Children's earliest successes have been documented on measures that allow *approximate* responses, such as matching multidigit number names to written digits, or correctly judging that number names like "two hundred fifty-three" map onto the written numeral "253" and not "532." Preschool children can also judge relative magnitudes given the written digits (e.g., knowing that "253" is more than "165"). Such measures are considered approximate because they do not require a precise interpretation of base-ten units and counts. Instead, children may correctly solve these approximate measures using various heuristics (e.g., knowing that numerals with more digits represent larger quantities) (*Authors*, 2022b; Mix et al., 2014). A precise interpretation requires an understanding of the syntax that combines places and the units counted. Eventually, children demonstrate competence on measures that directly query *syntactic* responses, such as counting sets of base-ten blocks or answering direct questions such as, “Which number has six tens?” (*Authors*, 2022b; Chan et al., 2014), usually after they have received formal instruction on multidigit numbers in school. Ultimately, this syntactic understanding of place value is recruited to carry out advanced arithmetic operations, including multidigit calculation (*Authors*, under review).

The emphasis in these prior studies has been children’s correct responses on measures that vary in their task demands. However, another source of information about children’s emerging base-ten concepts is the pattern of errors they exhibit on these measures. On measures that directly query syntactic understanding, for example, there is evidence that children generate responses that, while incorrect, reflect partial “not-quite-right” understandings of base-ten structure—what we will call *smart errors*. For example, Byrge et al. (2014) reported that when 4- to 6-year-olds are asked to write multidigit numbers, they sometimes add zeroes in a way that reflects the magnitudes of various base-ten units (e.g., 600405 for 645). There is little research examining the concurrent performance and longitudinal outcomes for children who exhibit smart errors compared to equally low-performing peers who do not. Further, it is currently unknown if smart errors emerge simultaneously across different base-ten tasks or emerge sequentially depending on specific task demands. In the present study, we take advantage of an existing longitudinal dataset to investigate the prevalence and developmental timing of these errors, and ask whether they are predictive of later mathematics outcomes.

Smart Errors in Place Value Learning

We focus on smart errors generated on two well-known measures of place value understanding—Transcoding and Base-Ten Counting. Correct performance on either measure strongly predicts later mathematics achievement (*Authors*, 2022a; *Authors,* under review; Chan et al., 2014; Moeller et al., 2011), and both measures are central organizers in children's emerging conceptual structures for place value understanding, compared to other place value measures, such as expanded notation (*Authors,* 2022b). Importantly for our purposes, both measures are open-ended, allowing children to reveal their thinking in a way that can be obscured in forced-choice measures. We also know from previous work that children exhibit characteristic errors on both Transcoding and Base-Ten Counting—errors that seem indicative of structural awareness even before children have completely worked out the multiplicative structure of place value symbols and acquired the necessary vocabulary to achieve correct performance (Byrge et al., 2014; Chan et al., 2017; Fuson et al., 1997; Power & Dal Martello, 1990; Vasilyeva et al., 2022). Below we describe each place value measure in detail, along with the smart errors that have been documented for them.

Transcoding

The Transcoding task measures the mapping of number names to written forms through reading and writing (Byrge et al., 2014; Deloche & Seron, 1982; Moeller et al., 2011; Power & Dal Martello, 1990; Zuber et al., 2009). Although prior research has documented both syntactic and lexical errors in reading multidigit numbers (Vasilyeva et al., 2022), we focus here on writing of multidigit numbers. For the writing task, children are given a verbal number name and asked to write the corresponding multidigit numeral (e.g., "How do you write six hundred twenty-five?"). As noted above, children sometimes make *expanded errors* on this task by adding zeros to indicate base-ten value rather than relying solely on the spatial position of the digits (e.g., 600405) (Byrge et al., 2014; Power & Dal Martello, 1990). Though incorrect, these errors clearly reflect base-ten structure and arguably improve the alignment between verbal number names and their written counterparts by explicitly writing digits that map onto the words for units, rather than implying these units by spatial position. Just as children sometimes introduce more predictable structure to natural language by, for example, overregularizing past tense verb forms (e.g., saying "runned" instead "ran") (Hudson Kam & Newport, 2009; Rumelhart & McClelland, 1987)—linguistic forms they have not observed in use but rather, have invented themselves—the presence of expanded errors may reflect children's attempts to make the conventional system for place value representation more predictable and transparent.

When Byrge et al. (2014) tracked the emergence of expanded errors, they found that children first exhibited such errors at 4 years of age, though most responses at this age were random digit strings or other marks. However, for 5- and 6-year-olds, expanded errors were the most frequent response (greater than the correct conventional form and all other errors), with the proportion of correct responses increasing over this age range. In the present study, we track the emergence of these errors in a larger, longitudinal dataset and ask whether children who exhibit expanded errors in kindergarten go on to have stronger mathematics outcomes in first and second grade.

Base-Ten Counting

In the Base-Ten Counting task (Chan et al., 2014, 2017; *Authors*, 2022a; *Authors*, 2022b), children are asked to count line drawings of base-ten blocks representing various quantities (see Figure 1). The task is designed to elicit counting by base-ten units (e.g., counting 143 "small squares" as "100-110-120-130-140-141-142-143"), but children typically approach the task using various strategies, including counting by ones, counting within each base-ten unit separately, or counting by base-ten units as in the above example (Chan et al., 2014; Fuson et al., 1997). Children's accuracy at Base-Ten Counting in kindergarten and first grade is a reliable predictor of later mathematics achievement, regardless of which strategy they use (*Authors*, 2022a; Chan et al., 2014). However, focusing only on accurate responses might miss an earlier stage of development during which children grasp the essential idea that when you shift from one base-ten unit to another (e.g., from hundreds to tens) you are counting something different, even if you lack the vocabulary to accurately count by these base-ten units. During testing, we informally observed that children sometimes exhibited this awareness by shifting the words they used while counting at each of these unit boundaries but using either incorrect or invented vocabulary to signify these shifts. For example, at the 100-to-10 unit shift in the number 230, children may correctly count the two 100-units by 100s but then incorrectly change their counting vocabulary when they start counting the three 10s but not to tens (e.g., two-hundred ten..) but to ones (e.g., "one hundred, two hundred, two hundred one, two hundred two").

Previous research has examined the distribution of specific error patterns in children's Base-Ten Counting (Chan et al., 2017), but it has focused on the misconceptions. Rather than further investigating these error types, we focus instead on a more basic question—do young children make errors that reflect an awareness of shifts from one base-ten unit to another, regardless of what vocabulary they use to signify the shift? Although past research has not documented such invented counts, it has demonstrated that younger children use less effective strategies than older children and also make more errors overall (Chan et al., 2014, 2017). As tested in the Base-Ten Counting task, lower achieving children also make more errors than higher achieving children—particularly more random errors (Chan et al., 2017). These performance differences are consequential. In one study, children's accuracy on Base-Ten Counting was the strongest predictor of later mathematics outcomes even after controlling for age, nonverbal intelligence, and several other multidigit numeracy skills, including Transcoding

(Chan et al., 2014). The present study investigates whether partial understanding of unit boundaries, though incorrect, may also prove to be a useful predictor.

[INSERT FIGURE 1 HERE]

The Developmental Contribution of Approximate Place Value Understanding

As noted above, children begin learning place value concepts with approximate or intuitive understandings (*Authors*, 2022a; *Authors*, 2022b; Byrge et al., 2014; Mix et al., 2014; Yuan et al., 2019). Previous work has demonstrated that kindergartners' performance on approximate tasks (i.e., transcoding (reading and writing number names); magnitude comparison (Which is larger? 119 or 191?); and number line estimation (marking where 43 goes on a 1-100 number line) is separable from their performance on tasks that require a more precise response (i.e., digit-place correspondence (e.g., Which number has two *tens*? 230, 120, or 542); base-ten counting (counting base-ten blocks), and expanded notation (Which of these add up to 83? 800 + 3 or 80 + 3) using both confirmatory factor analysis and community detection in a network analysis (*Authors*, 2022a; *Authors*, 2022b). Importantly, only accuracy on approximate tasks in kindergarten significantly predicted children's syntactic understanding of place value in first grade (*Authors*, 2022a) as well as predicting their multidigit calculation scores in second grade (*Authors*, under review). Kindergarten accuracy on syntactic skills did not predict either of these outcomes, suggesting that implicit, partial knowledge of multidigit number meanings provides a foundation for later explicit understanding of base-ten principles. Perhaps partial understandings revealed through smart errors steer children's attention and facilitate their discovery of the fundamental syntax, as exemplified in the broader context of learning (Gentner, 2010).

However, it is possible that partial knowledge of base-ten syntax, reflected in smart errors, would significantly predict both later precise syntactic skill and eventually, children's multidigit calculation skill. Indeed, there may be heterogeneity among children who succeed on approximate measures that is masked by an approach that relies exclusively on the analysis of correct performance. Specifically, some children may be correct on approximate measures based on rough estimates of relative quantity and very limited heuristics that signal associations with these estimates (e.g., 478 > 3 because lots is more than a little), whereas other children (or perhaps the same children in other contexts) are correct on approximate measures because they have enough sense of base-ten structure to make strong guesses (e.g., 478 > 3 because there's a one-one correspondence between the digits in the ones place for these two numerals, but nothing else to align in the other places). Correct performance on approximate measures does not tell us which reasoning children are using. We know from their failure to perform syntactic measures that they are not generating correct responses by fully unpacking base-ten structure (e.g., 478 > 3 because [(4 x 100) + (7 x 10) + (8 x 1)] > (3 x 1)). However, some children may be developing some ideas about these structural properties of multi-digit numbers. We hypothesize that smart errors, such as invented counts and expanded errors, may be more sensitive to the earliest emergence of these structural ideas than accuracy on either approximate or syntactic place value measures, and thus may reveal longitudinal relations between syntactic understanding and later place value and multidigit calculation skill that were missed by examining correct performance alone. Importantly, we think these smart errors may uniquely predict later place value and multidigit calculation learning above-and-beyond general cognitive ability. Thus, in our current analyses we control for children's performance on a matrix reasoning task.

Current Study

This study examined the development of two smart errors observed in place value learning—expanded errors on the Transcoding task and invented counts on the Base-Ten

Counting task—through secondary analysis of a longitudinal dataset that includes measures of children's place value understanding and multidigit calculation skill from kindergarten to second grade (*Authors*, 2022a; *Authors*, under review). The study addressed three specific questions:

1. Do both types of smart errors (expanded errors and invented counts) emerge concurrently? They might if both reflect the same emerging but partial understanding of base-ten structure. However, each error might emerge on its own time course if dependent on knowledge of different aspects of base-ten structure.
2. Is the presence of one or both smart errors in kindergarten associated with higher levels of place value understanding on other place value measures? In past research, young children exhibited earlier correct performance on approximate measures (*Authors*, 2022a) based on partial understanding. This same partial understanding may be evident in certain kinds of incorrect performance (i.e., smart errors) that can be observed on two open-ended measures, Transcoding and Base-Ten counting.
3. Do kindergarteners who exhibit smart errors go on to experience greater place value growth between kindergarten and first grade than those who do not? This pattern may be obtained if these partial understandings indicate a foundation for learning about place value in school.
4. Is evidence of either or both smart errors associated with higher levels of multidigit calculation skill in second grade? Previous research has already established that accuracy on place value measures in kindergarten and first grade is associated with better multidigit calculation skill in second grade (*Authors*, under review; Chan et al., 2014; Moeller et al., 2011). Here, we ask whether inaccurate responses that reflect awareness of base-ten structure are also strong predictors.

Method

Participants

Children were tested at three timepoints: spring of kindergarten, first grade, and second grade. At kindergarten testing, the sample consisted of 279 children (135 females; 144 males) with a mean age of 5.76 years (SD = 0.55). At first grade testing, the sample consisted of 231 children (117 females; 114 males) with a mean age of 7.15 years (SD = 0.37). At second-grade testing, the sample consisted of 197 children (97 females; 100 males) with a mean age of 8.14 years (SD = 0.36). Attrition from one grade level to the next occurred because children had moved and could not be located. A sensitivity analysis conducted in G*Power (Faul et al., 2009) indicated that a sample size of 197 would be adequate to detect a medium effect using a multiple regression model (i.e., Cohen's f^2 = .12, Cohen, 1988).

Children were recruited from four cities in the Midwestern and Mid-Atlantic regions of the United States: 40 children were from [*state blinded for review*], 186 children from [*state blinded for review*], and 53 children from [*state blinded for review*]. Families of 213 children provided written consent for their children's participation. For the remaining 66 children, school administrators requested an IRB-approved opt-out consenting process in which families were notified but only returned their consent forms to indicate exclusion from the study. None of the families opted for exclusion. Most families (54%) either did not return demographic information when given the demographic questionnaire or were not given a demographic questionnaire to complete because of the opt-out consent process. To estimate the missing demographic data for these families, we used school-wide information for 31% and 2017 neighborhood census data for the rest. Weighted sample descriptive statistics indicated that the sample was racially diverse

(41% Black, 38% White, 12% Latino, 8% Asian), and primarily middle socioeconomic status (average median family income range = $75,000 to $99,999).

Procedure and Materials

Testing sessions took place in a quiet area outside of the classroom and lasted approximately 60 minutes per child. All measures were administered individually in one of two random orders, counterbalanced across children. Reliabilities were calculated using Cronbach's alpha unless otherwise noted. Two measures had reliabilities below .70 (Which N Has? and Expanded Notation), which might suggest multidimensionality and weaken the strength and generalizability of the results in the current study. However, others have argued that such measures may be retained if they provide important content coverage (Schmitt, 1996). We thus kept both measures in our analysis because they provide specific content coverage needed for the goals of this manuscript; however, we cautiously interpret any results for them with these reliabilities in mind.

Place Value Skills

Six place value skills were assessed in kindergarten and first grade: three approximate place value tasks (Transcoding, Number Line Estimation, Magnitude Comparison) and three syntactic place value tasks (Base-Ten Counting, Which N has __?, Expanded Notation).

Transcoding (e.g., Byrge et al., 2014). Transcoding is the abilty to read and write numerals. For the reading assessment, children saw a stimulus number (e.g., "23") and said its name aloud while the experimenter recorded their response (e.g., "twenty-three"). For the writing assessment, children listened to the experimenter say a multidigit number name and were told to write down the numeral they heard. Both the reading and writing assessments were comprised of one 2-digit number; one 3-digit number; and one 4-digit number, for a total of twelve test trials across the two assessments. Trials were coded as either correct or incorrect (maximum possible score = 12). Partially correct responses were not counted as correct (e.g., reading the numeral 239 as two hundred three-nine); however, full credit was given for written responses that involved numeral reversals (e.g., writing 3 backwards) (α = .87). These total scores were included in the factor scores that were used as developmental outcomes in some analyses; however, it was only possible to observe expanded errors on the six written trials.

Coding for Expanded Errors. Following Byrge et al. (2014), children received a score of "0" for each written response where there was no expanded error and a score of "1" for each written response where an expanded error was made (maximum number of expanded errors = 6). Expanded errors were coded on responses that contained all the correct digits in order, but included additional numerals between them (i.e., either extra 0's or 1's). For example, writing the numeral "two thousand seven hundred forty-three" as "200743" would be considered an expanded error, as would "2000700403," "200070043," "2700403," and so on. All participant responses for the six items were first coded by the primary coder, and then agreement with a second coder was assessed on 20% of the items. Intercoder reliability was high with 97% agreement for kindergarten and 98% agreement for first grade. We then created a variable that divided the number of expanded errors by the number of incorrect trials to obtain a proportion score of the number of expanded errors produced on incorrect trials.

Base-Ten Counting (Chan et al., 2014). In this task, children counted various quantities represented with line drawings of base-ten blocks (see Figure 1A). Prior to the test trials, children were first provided with a short five-minute introduction to base-ten blocks. The experimenter displayed a 10s-block and demonstrated how ten of the "small squares" were combined to make a 10s-block, counting the individual squares by ones. Next, the experimenter introduced the 100's-block, first showing how 100 small squares came together to make the larger 100-block. Initially, the experimenter counted the small squares by ones, but then stopped, noting that "counting all those small squares would take a really long time," showing instead how ten 10's-blocks could make the larger 100-block, counting them by tens to illustrate. After this short introduction, children were shown a line drawing of a physical representation of a quantity (e.g., 13 represented with base-ten blocks) and told it was a picture of the same blocks. Children were then asked to tell the experimenter how many small squares were in the picture. There were 10 total trials: five trials with representations of two-digit quantities and five trials with representations of three-digit quantities. Children were permitted to count by ones on the first trial, but if they attempted to do so on the second trial (for which the target number was 42), they were allowed to finish, and then reminded that these blocks could also be counted by tens. Children were then allowed to count again, and the better of the two trials was scored. A similar prompt was given if children attempted to count the first trial with 100-blocks by ones. Notes on children's counts were recorded but their scores were the total correct in terms of identifying the number of small squares in each drawing (maximum possible score was 10) (α = .85).

Coding for Invented Counting Errors. Across the ten items, each base-ten unit shift (e.g., tens-bars shifting to ones-squares) was coded on incorrect trials to reflect whether or not children changed their counting strategy when the base-ten unit changed even when children produced incorrect counts. Clearly accurate counts change when base-ten units shift, but our aim was to determine whether children's inaccurate counts also reflected a shift at the base-ten unit boundaries. For example, if an item contained three tens-bars and four ones-squares ("34"), a child who lacked the vocabulary to count correctly might "count" the tens-bars with an invented system (e.g., 101, 102, 103) and then switch to counting the ones-squares by ones (e.g., 1, 2, 3, 4).

We also excluded two items that had the ones-squares not grouped together, but instead were mixed among the tens-bars because these items did not lend themselves as easily to counting by base-ten units (see Figure 1B for an example of an excluded item). Of the eight remaining items, four items consisted of only one shift in base-ten units (100-to-10 or 10-to-1) and four items consisted of two shifts in base-ten units (100-to-10 and 10-to-1). Thus, there were 12 shifts in representational units across the eight items. Each of the 12 shifts was coded for invented counts on incorrect trials by two independent coders for all participants. Intercoder reliability was high (Kappa = .91; McHugh, 2012); however, when there was a discrepancy, the first author discussed it with the two coders and all agreed upon the code. Because not all children attempted every item (e.g., some of the later, more difficult items were skipped), children's total number of invented counts was divided by the total number of shifts in representational units they attempted, which produced a proportion score of boundaries distinguished at changes in representational units.

"Which N has __ ?" This task was a multiple-choice adaptation of the digit correspondence task that has been used in previous research (Hanich et al., 2001; Kamii, 1989). Children were presented with three written numerals arranged in a horizontal line (e.g., 2, 20, and 10). The experimenter then asked the child to select the number that answered a place value question such as, "Which number has two tens?" After two practice trials, the six test trials included two trials each probing tens, hundreds, and thousands. The position of the correct response was counterbalanced across trials. Test trials were coded as either correct or incorrect (maximum possible score = 6, chance = 2) (α = .53).

Expanded notation. This commonly used multiple choice task asked children to match written numerals to their expanded notation forms (e.g., Mix et al., 2016). Children were shown a written numeral (e.g., 73) and asked to select the correct expanded version from among three options (e.g., 70+3, 700+3, or 7+3). The choices were arranged vertically on the right side of the page, and the target number was presented in a larger font on the left side. Before the test trials, the experimenter explained that the plus sign means combining two numbers. Then children were asked, “Which of these (pointing to the equations) adds up to be this number (pointing to the target)?” There were two practice trials with corrective feedback, followed by six test trials—two trials each probing two-digit, three-digit, and four-digit numbers. Test trials were coded as either correct or incorrect, for a maximum possible score of 6 (chance = 2) (α = .57).

Number line estimation (Siegler & Opfer, 2003). Children were given a blank 0-100 number line and told to indicate where a number (e.g., 3) should be located using a vertical mark. The target number was printed at the top of the page in the center. There was one practice trial and 15 test trials. The test trials were coded for percentage of absolute error (PAE) by measuring the distance from the hatch mark to the correct location and dividing by the scale (in this case, 100). The total score was the PAE averaged across the 15 test trials (range = 0 – 90% PAE, even-odd reliability: r = .76)

Magnitude comparison. The aim of this task was to indicate which of two written numerals represented the larger quantity (e.g., 461 vs. 614). The choice numerals were printed on opposite sides of an 8 x 11 inch sheet of paper and presented one by one by flipping pages in a binder. There were 25 trials comprised of one to four-digit numerals (trials adapted from Mix et al., 2014). Correct responses received one point, for a total possible of 25 (chance = 12) (α = .72).

Multidigit Calculation Skill

Calculation skill was assessed in second grade using the addition and subtraction subtests of the Comprehensive Mathematical Abilities Test (CMAT; Hresko et al., 2003). Test items were arranged in order of difficulty moving from whole number calculation to problems with decimals, fractions, and mixed numbers. Of the 48 items, 30 items used whole numbers, and of these, 24 items used multidigit numerals and 6 were a mix of single-digit only items and mixed multi-single digit items. Testing stopped when children answered three items in a row incorrectly. The CMAT outcome variable used in analyses was the combined age-based standardized score that was determined from each child’s addition and subtraction raw scores. Reliability from the norming sample was high for both subtests ($\alpha = .86$).

Matrix Reasoning

We used children’s scores on the Matrix Reasoning subtest from the *Wechsler Intelligence Scale for Children* – 5th edition (*WISC-V*; Wechsler, 2014) to estimate general cognitive ability. Matrix Reasoning was assessed at the second test session when children were in first grade. The subtest consisted of two practice items and 32 test items in which children choose a figure that completes a repeating pattern or visual analogy. Based on the *WISC-V* testing procedures, children completed items that increased in difficulty until they gave three consecutive incorrect responses. The age-based standardized score was used as a covariate in the current analyses. Reliability from the norming sample was high ($\alpha > .80$).

Analysis Plan

We first provide descriptive statistics (means, standard deviations, counts) to characterize the frequencies of expanded errors and invented counts in kindergarten and first grade. Recall that in Byrge et al.’s (2014) cross-sectional transcoding study, the frequency of expanded errors was greater at 5 years of age than 4 years of age and remained the most frequent response at ages 5 years and 6 years. To assess whether the same patterns were evident in a longitudinal study, we used *t*-tests to compare the frequencies of expanded errors at each age point (kindergarten and first grade), as well as providing a qualitative comparison of error rates with means reported by Byrge and colleagues.

We addressed our first main question of whether the two smart errors emerge at the same timepoint by dividing kindergartners into four groups based on whether they made no smart errors, only expanded errors, only invented counting errors, or both smart error types. Chi-square goodness of fit tests were conducted for each grade level to see if the distribution of children across the four groups differed significantly from chance. We also computed bivariate correlations to examine the associations between the two smart errors at each grade level.

Our second main question was whether the children who exhibited smart errors in kindergarten also had greater concurrent accuracy on approximate measures, syntactic measures, or both compared to children who did not exhibit smart errors. To evaluate this question, we conducted two analyses. First, for each of the three place value outcomes (general, syntactic, and approximate), a path analysis was conducted using the smart error continuous variables (i.e., proportion of smart errors produced on incorrect trials) as predictor variables of each of the three concurrent composite z-scored variables while controlling for performance on the Matrix Reasoning subtest. Second, because producing even one smart error (of either type) may be an indicator of children's partial, underlying place value understanding, we took a categorical approach. Specifically, the error grouping described above was used as the independent variable in an analysis of covariance (ANCOVA) that compared the mean performance of the four groups on approximate place value measures, syntactic place value measures, and multidigit calculation measures while controlling for Matrix Reasoning scores. For each of the three outcomes, we computed *z*-scores for each of the tasks, collapsed across time, and then computed kindergarten (and first grade) composite variables. Bonferroni-adjusted pairwise comparisons were used to examine specific subgroup differences for all ANCOVA analyses, unless otherwise noted.

To address our third main question of whether kindergarteners who made smart errors showed more growth in place value understanding by first grade than those who did not, we repeated the path analyses and ANCOVAs described for the concurrent analyses but used difference scores derived by subtracting children's kindergarten composite scores (approximate, syntactic, multidigit calculation) from their first-grade composite scores.

Our fourth main question was whether children who exhibited smart errors in kindergarten went on to have better multidigit calculation in second grade. To address this question, we used the same analytic approach as the previous analyses (i.e., path analysis and ANCOVA) using a z-scored composite variable of children's performance on multidigit addition and subtraction items in second grade.

Lastly, we would like to emphasize that our current approach of computing the proportion of smart errors produced on incorrect trials is an attempt to reduce the confound of task accuracies in the production of these smart errors. However, there may be children who were mostly correct on the tasks and produced only a few smart errors, so their proportion of errors on incorrect trials could be artificially high. Thus, the relations between the production of these smart errors and overall task performance may be misleading. Although using proportion of smart errors produced on incorrect trials allows us to assess the role of smart errors more stringently in the growth of children's place value understanding above-and-beyond task accuracy, we took a supplemental approach to address the previously-discussed potential issue.

In this secondary analysis, we removed 19 children from our sample who had above 80% accuracy on the Base-Ten Counting and conducted all analyses again. All results were the same as when we used all children in our sample. As such, the results reported here are with the entire sample.

Results

Descriptive Statistics

Expanded Errors (Transcoding)

As shown in Table 1a, expanded errors on the Transcoding measure were frequent, but correct answers were the most frequent response ($M_{\text{proportion correct}} = .53$, $SD = .29$). This finding contrasts with the findings reported by Byrge et al. (2014), in that children in their sample were incorrect on more trials. However, as in Byrge et al. (2014), the overall frequency of expanded errors on incorrect trials peaked in kindergarten ($M = .27$, $SD = .23$) and stayed relatively similar in first grade ($M = .25$, $SD = .26$), $t(130) = 0.80$, $p =.424$. Because children in the current study were slightly older by grade-level compared to Byrge et al. (i.e., kindergarteners in the current study were 6 years of age whereas kindergartners in Byrge et al. were mostly 4- and 5-year-olds), we also compared the frequencies based on age in years instead of grade. The percentage of 6-year-olds who produced at least one expanded error (65.6% on incorrect trials; 60% across all trials) was nearly identical to the percentage reported by Byrge et al. for the same age group (61%), thus replicating this previously reported finding.

In addition to children's overall error rate, we analyzed children's performance based on the magnitude of the numbers being requested (2-digit, 3-digit, 4-digit), to see whether expanded errors peaked and decreased at different ages depending on the size of the numbers that were queried. In kindergarten, there was a marginally significant difference between 2- and 3-digit trials in that children made slightly more expanded errors on 3-digit trials compared to 2-digit trials, $t(42) = 1.95$, $p = .054$, Cohen's $d = .30$. Additionally, children made significantly more expanded errors on 3-digit trials compared to 4-digit trials, $t(167) = 3.68$, $p < .001$, Cohen's $d = .28$. In first grade, children also made significantly more errors on 3-digit trials than 4-digit trials, $t(65) = 2.20$, $p < .031$, Cohen's $d = .27$. There was no significant difference in the frequency of errors on 2- versus 3-digit trials, $p = .477$.

[INSERT TABLE 1A HERE]

Invented Counting (Base-Ten Counting)

As shown in Table 1b, the number of invented counts increased from kindergarten to first grade, regardless of base-ten magnitude (see Table 1b for *t*-test statistics). Indeed, most children in first grade (60%) exhibited a high proportion of invented counts (> 80%). In terms of base-ten magnitude, both kindergartners and first graders exhibited more invented counts moving from hundreds-to-tens than from tens-to-ones [Kindergarten: $t(178)=3.40$, $p < .001$, Cohen's $d = .25$; First Grade: $t(194)=2.22$, $p = .028$, Cohen's $d = .25$].

[INSERT TABLE 1B HERE]

Developmental Relations of the Two Smart Error Types

The first question we addressed was whether children who made expanded errors also exhibited invented counts at the same time. As noted above, we divided children into four groups: (1) those who made neither error ($n = 34$; 14%); (2) those who made at least one expanded error only ($n = 91$; 38%); (3) those who produced at least one invented count error only ($n = 49$; 21%); and (4) those who produced both error types at least once ($n = 63$; 27%). The distribution was significantly skewed toward children exhibiting only expanded errors at this age, χ^2 $(3) = 29.79$, $p < .001$. Collapsing the two expanded error groups (i.e., expanded errors only plus both smart errors), further revealed that most children (*n* = 154; 65%) made at least one expanded error. Alternatively, few children made neither error (14%), casting doubt on the notion that kindergarteners successfully performed approximate measures without at least a partial understanding of base-ten structure. Rather, it appears that the vast majority of children displayed at least partial, not-quite-right understanding and this partial knowledge likely supported their performance on approximate measures, despite weaker performance on syntactic measures. In first grade, only 125 children made at least one error, with their group membership distributed as follows: (1) neither error (*n* = 6; 5%); (2) at least one expanded error only (*n* = 17; 14%); (3) at least one invented count error only (*n* = 42; 33%); and (4) both error types (*n* = 60; 48%). Unlike the distribution observed in kindergarten that was skewed toward expanded errors, the first-grade distribution was significantly skewed toward children exhibiting both smart error types, χ^2 (3) = 57.05, $p < .001$. It is also noteworthy that most children who answered incorrectly on at least one trial produced at least one invented count (81%), whereas in kindergarten, the majority of children produced at least one expanded error.

The age-related shifts in error distributions described above suggest that expanded errors on the Transcoding measure might peak earlier than invented counts in the Base-Ten Counting measure, in an overlapping waves pattern (Siegler, 1996). To evaluate this hypothesis, we tested the associations between the two error types at each age point using bivariate correlations. In kindergarten, the association was not significant ($r = -.11$, $p = .103$), but in first grade, a negative correlation approached significance ($r = -.16$, $p = .071$) suggesting that children who made more invented counting errors produced fewer expanded errors.

There also was a significant positive association between kindergartners' Transcoding *accuracy* and the frequency of invented counts in the Base-Ten Counting task, $r = .38$, $p < .001$, indicating that as children began to generate correct Transcoding responses, they also generated more smart errors on Base-Ten Counting. This evidence suggests that expanded errors peak in kindergarten, and were gradually replaced with correct performance in first grade (as Byrge et al. 2014 reported), whereas invented counting errors on Base-Ten Counting may lag behind, peaking in first grade based on the current dataset, and likely recapitulating the same pattern of decreasing errors as accuracy on Base-Ten Counting increases later in development.

Predicting Concurrent Place Value Knowledge

We next assessed whether the production of either (or both) smart errors differentially predicted children's concurrent understanding of place value above-and-beyond task accuracies. We were particularly interested in whether either the two smart error types differed in their prediction of approximate place value understanding as such a difference might signal heterogeneity in the way children approached these less precise measures—heterogeneity that yielded similar predictability of performance on approximate measures but different predictability of performance on syntactic measures (or on the longitudinal relations we evaluate in later sections).

General Place Value Understanding

To find out, we first conducted a path analysis with kindergartners' performance on the six place value measures and the two smart error frequencies as predictors of a composite variable for the concurrent six tasks' z-scores as described in the Analysis Plan section (while controlling for matrix reasoning ability). Results suggest that only invented counting errors on the Base-Ten Counting task was a significant predictor of general place value understanding, $\beta = .07$, $p < .001$, above-and-beyond task accuracies. In other words, kindergartners who produced invented counting errors—not expanded errors—on incorrect trials had better overall concurrent place value understanding than children who did not produce invented counting errors on incorrect trials.

Although the above path analysis treated the frequency of these two error types as continuous variables (i.e., how many errors are produced), the presence of even a single error token may indicate partial place value understanding. Thus, taking a categorical approach, we conducted an analysis of covariance (ANCOVA) comparing the overall place value performance of kindergarteners in all four groups based on the error patterns they exhibited. Using the same composite variable derived from performance on the six place value tasks (converted to z-scores), the ANCOVA revealed a significant group difference, $F(3,191) = 5.88$, $p < .001$, $\eta_p^2 = .08$. Bonferroni pairwise comparisons indicated this main effect was due to significantly higher place value scores for children in the invented counting error-only group compared to children in any of the other three groups, p's < .033[1].

Syntactic Place Value Understanding

We next repeated the analyses described above to see whether the presence of smart errors predicted performance on syntactic measures alone, using a composite of the children's scores on Expanded Notation, Base-Ten Counting, and Which N Has ___? with Matrix Reasoning controlled. The same pattern was obtained in that only the frequency of invented counting errors on the Base-Ten Counting task was a significant predictor of syntactic place value understanding, $\beta = .12$, $p < .001$, above-and-beyond task accuracies. In the ANCOVA based on children's smart error grouping with their syntactic composite variable as the outcome, a significant group difference was obtained, $F(3,191) = 4.06$, $p = .008$, $\eta_p^2 = .06$, such that children in the invented counting error-only group had higher syntactic scores than children in

[1] The pairwise comparison results are the same when using Tukey's Honestly Significant Difference test.

the expanded errors-only group, $p = .005$. Thus, results from both the path analysis and ANCOVA suggest that making at least one invented counting error is an indicator of children's syntactic place value understanding given that children who produced at least one of these errors had higher syntactic place value scores than those who made expanded-errors only. Moreover, the more invented counting errors children produced on incorrect trials, the higher their syntactic scores.

Approximate Place Value Understanding

We next repeated the path analysis using a composite of the children's scores on Magnitude Comparison, Number Line Estimation, and Transcoding, controlling for Matrix Reasoning scores; however, the model did not converge due to possible overfitting, or multicollinearity with the covariate of matrix reasoning. In the ANCOVA using the approximate composite variable as the outcome, a significant group difference was obtained, $F(3,191) = 4.79$, $p = .003$, $\eta_p^2 = .07$, such that children in the invented counting error-only group had higher approximate place value scores than either children in the expanded errors-only group, $p = .008$, or those who produced both errors, $p = .004$. Thus, these results suggest that making at least one invented counting error is an indicator of children's approximate place value understanding given that children who produced at least one of these errors had higher approximate place value scores than those who made both types of errors or expanded-errors only.

Taken together, the results indicate that children who exhibited invented counting errors on the Base-Ten Counting task—not Transcoding expanded errors—associated with overall concurrent place value understanding, and specifically more precise syntactic understanding. In terms of error type groupings, the invented counting-errors group tended to outperform other groups in general, syntactic, and approximate place value understanding.

Predicting Growth in Place Value Knowledge

We next asked whether the presence of smart errors differentially predicted growth in children's general place value understanding, syntactic place value understanding, or approximate place value understanding, above-and-beyond task accuracies. A path analysis using kindergartners' z-scored composite place-value performance and the two smart error frequency variables as predictors of children's later first grade z-scored composite place-value performance, while controlling for Matrix Reasoning ability, indicated that neither type of smart error committed in kindergarten predicted growth in place value understanding (p's > .078). The same was true when a composite of children's syntactic place value performance was used as the outcome measure (p's > .119), but when a composite of children's approximate place value performance was the outcome measure, Transcoding expanded errors rate was a significant predictor of growth, $\beta = .12$, $p = .037$.

When we conducted one-way ANCOVAs using children's kindergarten error type grouping as a between-subjects variable and their difference scores derived from the z-scored place value composite variables described previously as dependent variables, several significant differences emerged. Specifically, there was a significant effect of group on children's overall place value growth, $F(3,191) = 3.25$, $p = .023$, $\eta_p^2 = .05$. Bonferroni-adjusted pairwise comparisons, initially suggested there were no significant group differences; however, Fisher's Least Significant Difference (LSD) test indicated greater growth for children who produced expanded errors-only, $p = .016$, and both types of errors, $p = .017$, in comparison to those who produced invented counting errors-only. There also was a significant effect of group, $F(3,191) = 3.16$, $p = .026$, $\eta_p^2 = .05$ when only the approximate place value tasks were considered as outcome measures, with Bonferroni-adjusted pairwise comparisons again indicating no significant group differences, but Fisher's LSD test, indicating greater growth for children who produced expanded errors-only, $p = .011$, and both types of errors, $p = .023$, in comparison to children who produced invented counting errors-only. Only the ANCOVA with performance on syntactic tasks yielded no significant group effects ($p = .134$). Thus, the presence of Transcoding expanded errors—but not invented counting errors—predicted residualized growth in general and approximate place value understanding.

Smart Errors as Predictors of Multidigit Calculation Skill

Our final research question was whether the presence of smart errors in kindergarten predicted a more distal outcome measure—multidigit calculation skill in second grade. We conducted a path analysis similar to the previous path analyses and the results suggest that neither Transcoding expanded errors nor invented counts significantly predicted later multidigit calculation, p's > .132. When we conducted a one-way ANCOVA using the error type grouping as a between-subjects variable and mean multidigit calculation performance as the outcome, there was a significant effect of group, $F(3,162) = 4.01$, $p = .009$, $\eta_p^2 = .07$. Bonferroni-adjusted pairwise comparisons suggest that invented counting errors-only group had higher multidigit calculation performance compared to those who made neither error, $p =. 007$. See Figure 2 for a bar graph.

Overall, these results suggest that kindergartners who noticed the shift in base-ten units as they are counting objects in base-ten groupings—even if they do not use the correct vocabulary to label these shifts—have achieved an important insight into base-ten structure, which paves the way for superior multidigit calculation skill in second grade. This error alone is highly predictive, perhaps because producing only this type of smart error represents relatively advanced place value concept development in the overlapping waves pattern that was revealed above.

[INSERT FIGURE 2 HERE]

Discussion

This longitudinal study examined the development of place value concepts in kindergarteners and first graders in the United States. We specifically investigated the emergence of two smart errors—invented counting in the Base-Ten Counting task and expanded errors in the Transcoding task. Both responses were incorrect, but reflected more understanding of base-ten structure than random guesses. We found that the two errors emerged in an overlapping waves pattern, with expanded errors appearing first and peaking during kindergarten, before beginning to wane as invented counting errors increased. We also found that kindergarteners who made invented counting errors, but not expanded errors, exhibited stronger overall concurrent place value understanding—whether assessed by approximate measures, syntactic measures, or both—than children in the other three groups (expanded errors alone, invented counting + expanded errors, and neither error). Moreover, kindergarten children who produced both types of errors as well as expanded errors alone increased their general, specifically approximate, place value understanding in first grade more than children who produced invented counting errors alone. However, kindergarten children who made only invented counting errors had stronger multidigit calculation skill the following year in second grade. Thus, these smart errors appear to signal early partial knowledge that supports subsequent growth in place value understanding and eventual multidigit calculation skills.

The emergence of smart errors. Both smart errors were evident in kindergarten and first grade. Indeed, only 5% of children at the first-grade time point exhibited neither of these errors.

Both errors were based on approximations of the correct symbolic forms for base-ten representations in English. Invented counting errors were approximations of the shift in count words that must happen at the boundaries of base-ten units, but used the wrong words to signify this shift. Expanded errors were approximations of the way individual digits are used to represent base-ten units in writing, but used extra zeroes or ones to indicate these unit shifts rather than relying on spatial position alone. Thus, both errors might be said to tap into an awareness of base-ten units and best guesses as to how to represent these units symbolically.

However, there are also important differences between the two error types that might explain the overlapping waves pattern we observed. Invented counting errors arise during attempts to map verbal counting onto physical object groupings, whereas expanded errors arise during attempts to map verbal counting onto written numerals. Children encounter written multidigit numerals in school and in daily life. Indeed, they begin noticing patterns in written multidigit numerals in early childhood (Mix et al., 2014). Children also have likely practiced writing numerals in school by this age, though probably not for the magnitudes we queried, or in the same way (i.e., writing from dictation). In contrast, the Base-Ten Counting task is a novel activity used in experiments which children almost surely do not encounter at home, and which teachers are also unlikely to introduce in school. Perhaps these differences explain why expanded errors appeared earlier and peaked earlier than invented counting errors.

Another possible explanation for the observed overlapping waves pattern is that the mapping from words to written symbols may be inherently easier to perform than the mapping from words to physical quantities. This assertion makes sense because children learn to name the individual numerals (0-9) in early childhood, around the same time they are learning to count (Geary & vanMarle, 2016; Mix, 2009; Purpura et al., 2013), so it may be a smaller leap to incorporate base-ten units into these number names and symbols (e.g., three-hundred-two as 302 vs. three-zero-two), than it is to map the same count words and units onto base-ten blocks—objects that are introduced in school and therefore, have not been associated with count words for several years prior to school entry.

A third possibility is that children benefit from their earlier foothold with Transcoding (reading or writing multidigit numerals) in a causal way that supports better performance on Base-Ten Counting. Perhaps the mapping between words and written symbols highlights a common structure that prepares children to seek this structure in the world (e.g., physical instantiations). Armed with this first mapping, they may be more likely to seek these referents and to be more successful in performing word-to-object mappings when they do. Relatedly, correlational evidence suggests performance on Base-Ten Counting and Transcoding tasks are more strongly associated with each other than they are to other measures of place value knowledge, such as symbolic magnitude comparison (*Authors,* 2022b). Further research that uses training on Transcoding to assess its influence on Base-Ten Counting would be helpful to test this hypothesis directly.

Smart errors relate to conceptual growth. Children who made only invented counting errors in kindergarten had better place value understanding when measured concurrently compared to children who also (or either) produced transcoding expanded errors or no errors at all. This was true for both their approximate and syntactic place value understanding. When examining growth in place value skill from kindergarten to first grade, children who made transcoding expanded errors—either alone or in conjunction with invented counting errors—had significantly greater growth compared to children who made neither error and children who made only invented counting errors. When directly comparing how children with invented counting strategies alone compared to all the other children in their place value growth (i.e., collapsed into one group), the results were opposite to what we hypothesized: Children with invented counting strategies alone had less growth compared to all other children. However, as previously mentioned, this is most likely due to having less room for growth because these same children also had stronger place value skill at kindergarten.

When considering multidigit calculation skill two years later, kindergartens who exhibited invented counting errors alone had significantly better multidigit calculation skill in second grade compared to other children. In contrast, children who did not exhibit these errors were less prepared to learn about place value and had worse future outcomes. At a minimum, this finding means that these invented counting errors are useful indicators of children's developmental status when it comes to place value understanding, and their preparation to benefit from further mathematics instruction moving forward. Rather than simply being another way to be incorrect, these errors reflect partial knowledge that is strongly associated with eventual mathematics outcomes.

An intriguing question is whether these error patterns are a necessary stage of development that, in addition to signaling emerging competence, also contributes to place value growth in meaningful ways. In prior research with the same dataset, we found that the skills measured by both the Transcoding and Base-Ten Counting tasks were potent organizers of children's emerging place value understanding in kindergarten and first grade and acted as central hubs within larger networks of place value skill at both age levels (*Authors,* 2022b). These skills seem to anchor children's place value skills in a way that other skills do not. It is interesting to ask, then, whether this anchoring happens as a result of mastering the two skills, or rather, because of the nature of the mappings themselves and thus, may begin earlier, as children gain partial understanding of them.

Another way that these smart errors may serve as an important developmental stepping stone is that the solutions children reach, though incorrect, may reduce the degrees of freedom enough to help them gain traction in reasoning about other aspects of the place value system. These errors reflect a correct understanding of the underlying structure, but an incorrect understanding of how to express it using conventional notation and count words. It is likely that the discovery of this underlying structure is as powerful, if not more powerful, a driver of subsequent place value understanding than full mastery of these tasks. Future research that separates children's smart errors from random errors in a network analysis might help to distinguish between these accounts.

Although the present effects involving expanded errors were not as strong as those involving invented counting, this may be due to the age groups we tested. The two error types did not emerge simultaneously but rather, emerged in an overlapping waves pattern with invented counting errors appearing after expanded errors. If we had measured initial place value understanding in preschool rather than kindergarten, it is likely that only expanded errors would have been apparent, and perhaps, the presence of these errors at that age would also have been significant indicators of later place value understanding and multidigit calculation. Further longitudinal research starting with younger children is needed to directly test this hypothesis.

Educational implications. The present findings reveal emergent understandings that teachers could leverage instructionally in several ways. First, by recognizing that these smart errors signal an awareness of base-ten structure, teachers could build on them by highlighting the errors themselves as ways to represent base-ten structure (e.g., "Some children write the number, *two hundred thirty-six* like this: 200306. How is this right? How is it wrong?"). By pointing out that these errors reflect the right structure (albeit in an unconventional way), teachers can both help the children who are making the error translate their ideas into conventional notation, while also helping children who are not yet making the errors profit through exposure to these more transparent representations of structure than the conventional symbols offer.

A second educational application is to use these errors as part of the diagnostic screening process to identify children in need of additional instruction. The present results point out clearly that errors on place value measures do not all signify the same level of understanding. Rather, the presence of smart errors indicates a more advanced level of understanding that might require a different instructional approach and lower dosage than other error types. For example, remediating children who exhibit smart errors might simply require an explicit conversation acknowledging how they are correct about the base-ten structure but also pointing out how we represent this structure in English. In contrast, children who do not yet exhibit these errors may need to build up to discovering the underlying structure itself, through more exposure to the symbols and various pairwise mappings (e.g., words to objects, words to numerals, etc.). Such children should be encouraged if they begin making either invented counting or expanded errors as these errors reflect advanced understanding and perhaps, conceptual stepping stones as well.

Conclusion

In sum, the smart errors we targeted in this study proved to be common in the early grades, and reflective of advanced place value understanding compared to other error types. Indeed, children who exhibited smart errors in kindergarten also exhibited better concurrent place value performance, and stronger multidigit calculation skill two years later in second grade. As such, these errors may serve as helpful markers when screening young children for later mathematics difficulties, and determining the best instructional approach. Although our findings are consistent with a causal link between this early, partial understanding and later multidigit number competence, further research is needed to demonstrate this causality, identify its mechanisms, and determine how best to leverage these mechanisms instructionally. For example, given the overlapping waves pattern revealed in the current study, future research might examine if instruction based on reading and writing numbers (Transcoding) leads to better place value understanding than instruction based on physical representations (Base-Ten Counting). However, the present results provide strong evidence that these errors reflect structural knowledge of place value that is, while flawed, a potentially important developmental contributor.

Acknowledgements
Removed for peer review

Data Availability Statement
Deidentified data are available via [*removed for peer review*].

Declaration of Interest Statement
The authors have no conflicts of interest to declare.

References

Byrge, L., Smith, L. B., & Mix, K. S. (2014). Beginnings of place value: How preschoolers write three-digit numbers. *Child Development*, *85*(2), 437–443. https://doi.org/10.1111/cdev.12162

Carpenter, T. P., Fennema, E., & Franke, M. L. (1996). Cognitively guided instruction: A knowledge base for reform in primary mathematics instruction. *The Elementary School Journal*, *97*(1), 3–20. https://doi.org/10.1086/461846

Chan, W. W. L., Au, T. K., Lau, N. T. T., & Tang, J. (2017). Counting errors as a window onto children's place-value concept. *Contemporary Educational Psychology*, *51*, 123–130. https://doi.org/10.1016/j.cedpsych.2017.07.001

Chan, W. W. L., Au, T. K., & Tang, J. (2014). Strategic counting: A novel assessment of place-value understanding. *Learning and Instruction*, *29*, 78–94. https://doi.org/10.1016/j.learninstruc.2013.09.001

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). Lawrence Erlbaum Associates.

Deloche, G., & Seron, X. (1982). From one to 1: An analysis of a transcoding process by means of neuropsychological data. *Cognition*, *12*(2), 119–149. https://doi.org/10.1016/0010-0277(82)90009-9

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Fuson, K. C. (1990). Conceptual structures for multiunit numbers: Implications for learning and teaching multidigit addition, subtraction, and place value. *Cognition and Instruction*, *7*(4), 343–403. https://doi.org/10.1207/s1532690xci0704_4

Fuson, K. C., Smith, S. T., & Lo Cicero, A. M. (1997). Supporting Latino first graders' ten-structured thinking in urban classrooms. *Journal for Research in Mathematics Education*, *28*(6), 738–766. https://doi.org/10.2307/749640

Geary, D. C., & vanMarle, K. (2016). Young children's core symbolic and nonsymbolic quantitative knowledge in the prediction of later mathematics achievement. *Developmental Psychology*, *52*(12), 2130–2144. https://doi.org/10.1037/dev0000214

Gentner, D. (2010). Bootstrapping the mind: Analogical processes and symbol systems. *Cognitive Science, 34*(5), 752-775. https://dx.doi.org/10.1111/j.1551-6709.2010.01114.x

Gervasoni, A., Hadden, T., & Turkenburg, K. (2007). Exploring the number knowledge of children to inform the development of a professional learning plan for teachers in the Ballarat Diocese as a means of building community capacity. In J. Watson & K. Beswick (Eds.), *Mathematics: Essential Research, Essential Practice* (Vol. 1, pp. 305–314). MERGA Inc.

Hanich, L. B., Jordan, N. C., Kaplan, D., & Dick, J. (2001). Performance across different areas of mathematical cognition in children with learning difficulties. *Journal of Educational Psychology*, *93*(3), 615–626. https://doi.org/10.1037/0022-0663.93.3.615

Hiebert, J., & Wearne, D. (1996). Instruction, understanding, and skill in multidigit addition and subtraction. *Cognition and Instruction*, *14*(3), 251–283. JSTOR. https://doi.org/10.1207/s1532690xci1403_1

Hresko, W. P., Schlieve, P. L., Herron, S. R., Swain, C., & Sherbenou, R. J. (2003). *Comprehensive Mathematical Abilities Test (CMAT)*. PRO-ED.

Hudson Kam, C. L., & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology*, *59*(1), 30–66. https://doi.org/10.1016/j.cogpsych.2009.01.001

Kamii, C. (1986). Place value: An explanation of its difficulty and educational implications for the primary grades. *Journal of Research in Childhood Education*, *1*(2), 75–86. https://doi.org/10.1080/02568548609594909

Kamii, C. (1989). *Young children continue to reinvent arithmetic, 2nd grade.* Teachers College Press.

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, *22*(3), 276–282.

Mix, K. S. (2009). How Spencer made number: First uses of the number words. *Journal of Experimental Child Psychology*, *102*(4), 427–444. https://doi.org/10.1016/j.jecp.2008.11.003

Mix, K. S., Levine, S. C., Cheng, Y.-L., Young, C., Hambrick, D. Z., Ping, R., & Konstantopoulos, S. (2016). Separate but correlated: The latent structure of space and mathematics across development. *Journal of Experimental Psychology: General*, *145*(9), 1206–1227. https://doi.org/10.1037/xge0000182

Mix, K. S., Prather, R. W., Smith, L. B., & Stockton, J. D. (2014). Young children's interpretation of multidigit number names: From emerging competence to mastery. *Child Development*, *85*(3), 1306–1319. https://doi.org/10.1111/cdev.12197

Mix, K. S., Smith, L. B., & Crespo, S. (2019). Leveraging relational learning mechanisms to improve place value instruction. In A. Norton & M. W. Alibali (Eds.), *Constructing Number: Merging Perspectives from Psychology and Mathematics Education* (pp. 87–121). Springer International Publishing. https://doi.org/10.1007/978-3-030-00491-0_5

Moeller, K., Pixner, S., Zuber, J., Kaufmann, L., & Nuerk, H.-C. (2011). Early place-value understanding as a precursor for later arithmetic performance—A longitudinal study on numerical development. *Research in Developmental Disabilities*, *32*(5), 1837–1851. https://doi.org/10.1016/j.ridd.2011.03.012

Power, R. J. D., & Dal Martello, M. F. (1990). The dictation of Italian numerals. *Language and Cognitive Processes*, *5*(3), 237–254. https://doi.org/10.1080/01690969008402106

Purpura, D. J., Baroody, A. J., & Lonigan, C. J. (2013). The transition from informal to formal mathematical knowledge: Mediation by numeral knowledge. *Journal of Educational Psychology*, *105*(2), 453–464. https://doi.org/10.1037/a0031753

Rumelhart, D. E., & McClelland, J. L. (1987). Learning the past tenses of English verbs: Implicit rules or parallel distributed processing? In B. MacWhinney (Ed.), *Mechanisms of language aquisition* (pp. 195–248). Lawrence Erlbaum Associates, Inc.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, *8*(4), 350–353. https://doi.org/10.1037/1040-3590.8.4.350

Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking* (pp. viii, 278). Oxford University Press.

Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science (0956-7976)*, *14*(3), 237–243. https://doi.org/10.1111/1467-9280.02438

Vasilyeva, M., Laski, E., Veraksa, A., & Bukhalenkova, D. (2022). What children's number naming errors tell us about early understanding of multidigit numbers. *Journal of Experimental Child Psychology*, *224*, 105510. https://doi.org/10.1016/j.jecp.2022.105510

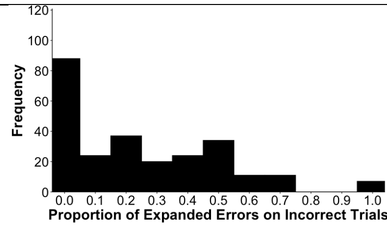
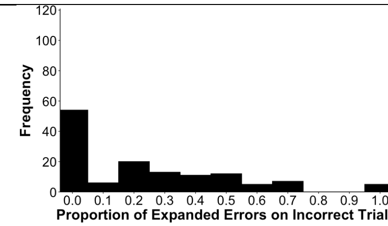
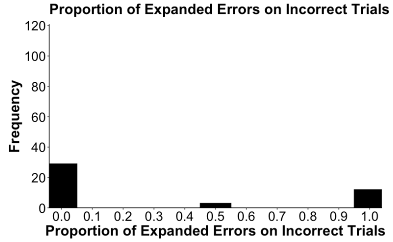
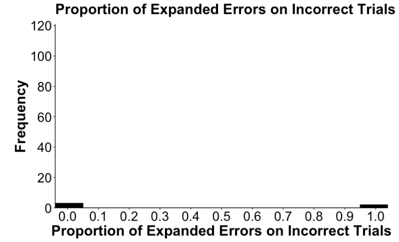
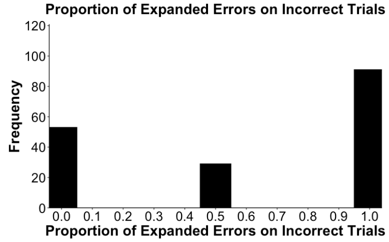
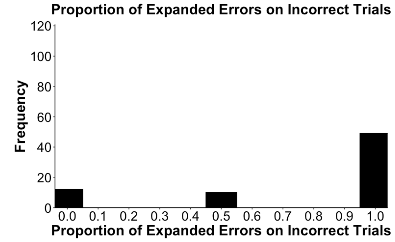
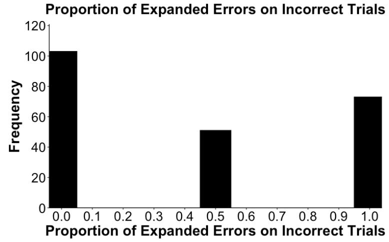
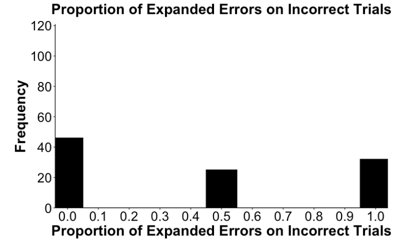
Wechsler, D. (2014). *WISC–V: Wechsler Intelligence Scale for Children integrated technical and interpretive manual* (5th ed.). Pearson.

Yuan, L., Prather, R. W., Mix, K. S., & Smith, L. B. (2019). Preschoolers and multi-digit numbers: A path to mathematics through the symbols themselves. *Cognition*, *189*, 89–104. https://doi.org/10.1016/j.cognition.2019.03.013

Zuber, J., Pixner, S., Moeller, K., & Nuerk, H.-C. (2009). On the language specificity of basic number processing: Transcoding in a language with inversion and its relation to working memory capacity. *Journal of Experimental Child Psychology*, *102*(1), 60–77. https://doi.org/10.1016/j.jecp.2008.04.003

Table 1a.

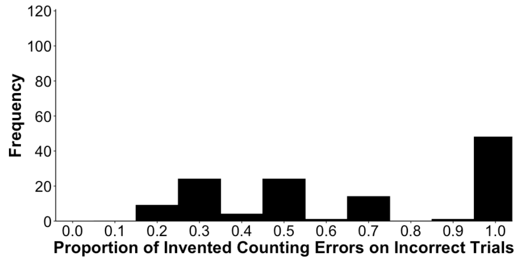
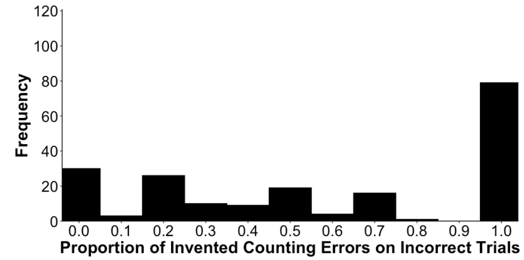
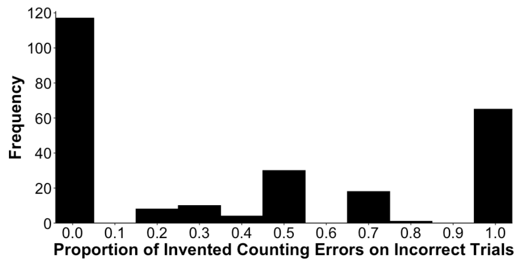
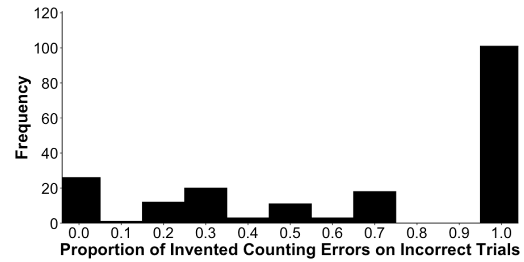
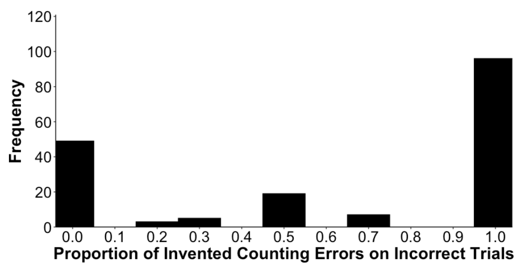
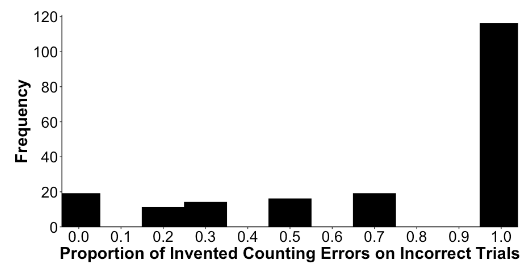
Descriptives of Proportion of Transcoding Expanded Errors Made Overall and by Grade and Trial Type

	Kindergarten		First Grade		
Trial Type	**Mean (SD)**	**Distribution**	**Mean (SD)**	**Distribution**	***t*-Test Results**
Overall	.25 (.25)		.25 (.26)		$t(130) = 0.80$, $p = .424$, Cohen's $d = .07$
2-digit	.31 (.45)		.40 (.55)		$t(2) = 0.38$, $p = .742$, Cohen's $d = .22$
3-digit	.61[a] (.44)		.76[a] (.39)		$t(62) = 1.40$, $p = .168$, Cohen's $d = .18$
4-digit	.43[b] (.44)		.43[b] (.43)		$t(100) = 1.26$, $p = .213$, Cohen's $d = .13$

Note: [a,b] Within grade comparisons: Kindergartners and first graders made significantly fewer expanded errors on 4-digit trials compared to 3-digit trials, *p's* < .031. However, children were at ceiling on performance on 2-digit trials (84% mean accuracy in kindergarten and 98% mean accuracy in first grade) so there was little to no opportunity to make expanded errors on these trials.

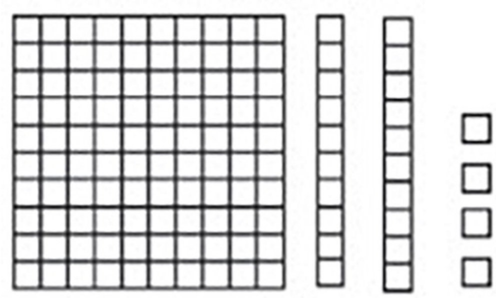
Table 1b.

Descriptives of Proportion of Unit Boundary Shifts Made Overall and by Grade and Boundary Type

	Kindergarten		First Grade		
Trial Type	**Mean (SD)**	**Distribution**	**Mean (SD)**	**Distribution**	***t*-Test Results**
Overall	.33 (.39)		.59 (.39)		$t(180) = 7.76$, $p < .001$, Cohen's $d = .58$
10-to-1 shifts	.40[a] (.42)		.68[a] (.38)		$t(179) = 7.99$, $p < .001$, Cohen's $d = .60$
100-to-10 shifts	.63[b] (.44)		.74[b] (.35)		$t(121) = 3.81$, $p < .001$, Cohen's $d = .35$

Note: [a,b] Children made significantly more 100-to-10 shifts than 10-to-1 shifts in both kindergarten, $p < .001$, and first grade, $p = .028$.

A.



B.

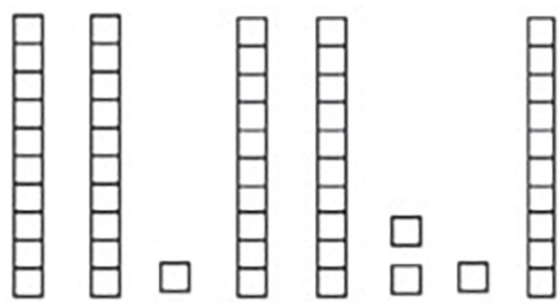


Figure 1. A. Example item from the Base-Ten Counting Task (Chan et al., 2014). B. An excluded trial from the coding of invented counts on the Base-Ten Counting Task.

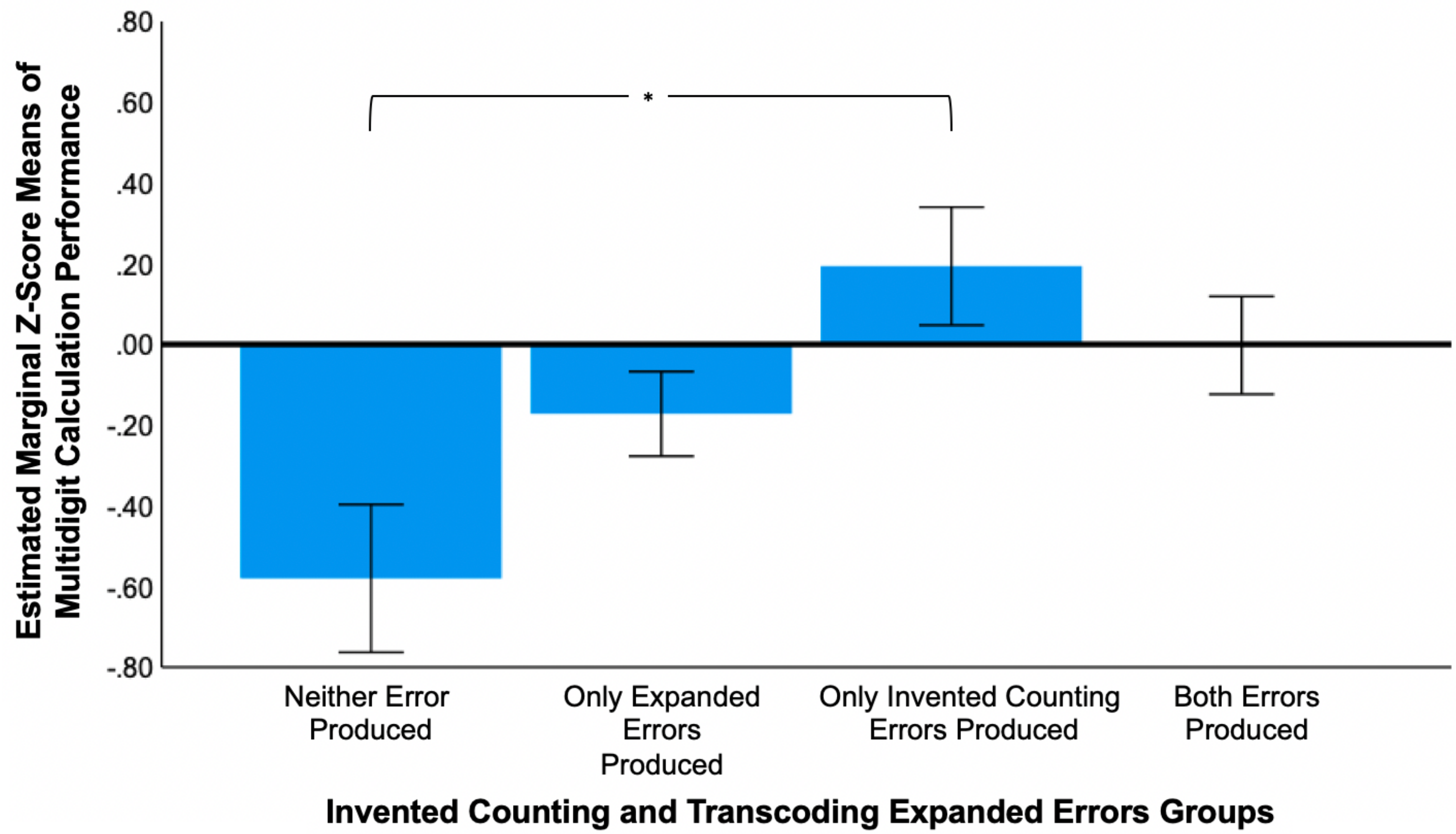


Figure 2. Mean multidigit calculation z-score (controlling for matrix reasoning scores) by kindergartners' invented counting and transcoding expanded errors groups. Error bars represent 1 standard error. *$p < .05$