Contemporary Educational Psychology 60 (2020) 101834

Contents lists available at ScienceDirect

Contemporary Educational Psychology

journal homepage: www.elsevier.com/locate/cedpsych

Strategy diversity in early mathematics classrooms☆

Douglas H. Clements[a,*], Denis Dumas[a], Yixiao Dong[a], Holland W. Banse[b], Julie Sarama[a], Crystal A. Day-Hess[a]

[a] University of Denver, United States
[b] University of Alabama, United States

ARTICLE INFO

Keywords:
Arithmetic
Classroom characteristics
Early childhood
Learning trajectories
Mathematics education
Strategies

ABSTRACT

Strategic processes are a form of procedural knowledge in which a child knows how to enact a given strategy that improves their capability in problem solving or learning. The solution strategies children use are critical components of their learning, especially in mathematics. Children vary substantially in their knowledge and use of different strategies, and much research has focused on intraindividual strategy variability. However, we do not know if classrooms that evince a broader variety of strategies across children are related to higher mathematics achievement. We investigated the diversity of arithmetical strategies within classrooms and examined the relations between strategy diversity and mathematical achievement as children moved from preschool to kindergarten and first grade. These analyses were applied to data from a large-scale experiment involving 1305 children from 42 schools and 106 classrooms. We created and applied a new method of measuring classroom strategy diversity and related this measure to children's concurrent and subsequent math achievement. We found that early strategy diversity was strongly related to achievement, but in subsequently, less diversity was so related. We compared these results to the predictions of three theoretical categories and found that our results mainly supported one.

Cognitive strategies are goal-directed and effortful procedures that children employ to aid in the regulation, execution, or evaluation of a problem or task (Alexander & Judy, 1988; Alexander, Grossnickle, Dumas, & Hattan, 2018). As such, strategic processes represent a form of procedural knowledge in which a child knows how to enact a given strategy that improves their capability in problem solving or learning (Dinsmore, 2017; Dumas, 2019). The solution strategies children use in mathematics are a particularly critical component of their learning in that domain (e.g., Biddlecomb & Carr, 2011; Carr, Jessup, & Fuller, 1999; Fennema et al., 1996; Pang & Kim, 2018; Sherin & Fuson, 2005; Wansart, 1990). Children's invented strategies may contribute to accuracy, problem-solving ability, base-ten number concepts, and flexibility in transferring knowledge in arithmetic (Carpenter, Franke, Jacobs, Fennema, & Empson, 1998).

When the same tasks are posed to multiple children, they often differ in the strategic procedures they employ to solve them (Magliano, Trabasso, & Graesser, 1999; Rhodes et al., 2019). Most research has focused on these as individual differences of strategy employment as well as on intraindividual strategy *variability* (although sometimes called "diversity" in the literature, we will use "variability" for intraindividual variety and reserve "diversity" for classroom, or interindividual, variety of strategies for clarity). Far less is known about the influence of classroom-level strategic diversity—the number of different strategies used within the classroom learning context.

Further, previous research has perennially considered the differential effectiveness of various strategic processes to improve student performance at the level of the individual student (Hickendorff, van Putten, Verhelst, & Heiser, 2010; Pressley & Harris, 2006). Specific instruction in mathematical strategies has been shown to be effective (Franke, Kazemi, & Battey, 2007; Jacobs & Empson, 2016), especially

☆ The original research was supported by the Institute of Education Sciences, U.S. Department of Education, through grants R305K05157 and R305A110188 and also by the National Science Foundation, through grants ESI-9730804 and REC-0228440 and the present analyses were supported in part by the supported in part by the Heising-Simons Foundation Grant #2015-156. The opinions expressed are those of the authors and do not represent views of the funders. A minor component of the intervention used in this research has been published by two of the authors, who thus could have a vested interest in the results. An external auditor oversaw the research design, data collection, and analyses of the original study (the present analyses have no implications for the curriculum). The authors wish to express appreciation to the original school districts, teachers, and children who participated in this research and to Megan Franke, who encouraged us to explore strategy diversity in classrooms.

* Corresponding author at: University of Denver, Denver, CO 80210, United States.
E-mail address: Douglas.Clements@du.edu (D.H. Clements).

https://doi.org/10.1016/j.cedpsych.2019.101834

Available online 23 December 2019
0361-476X/

for children with special needs (e.g., Fuchs et al., 2010; Naglieri & Johnson, 2000; Powell & Fuchs, 2015; Rockwell, Griffin, & Jones, 2011), although there are exceptions (e.g., Clearinghouse, 2017). However, these studies took quite different pedagogical approaches, from direct teaching of strategies to consistent evocation of a variety of strategies. Similarly, teachers receive conflicting messages regarding the benefits of eliciting diverse strategies and the different approaches to strategy instruction (Carpenter et al., 1998).

In summary, there is little research that investigates whether classrooms that evince a broader variety of strategies are related to higher mathematics achievement than classrooms with less diversity. That is, if children within a given classroom spontaneously enact strategies that differ from one another, would the level of strategic diversity be a positive or negative predictor of child performance in that classroom, and would that predictive relation vary depending on other factors (e.g., level of schooling, or individual characteristics)? In the current study, we investigate this question by measuring the diversity of mathematical strategies within classrooms and examine the relations between such strategy diversity and mathematical achievement as children moved from preschool to kindergarten and first grade. This age range is particularly appropriate because these are the years that arithmetic strategies are first developing for most children.

1. Background

We briefly discuss key findings from the literature because these findings contribute to the psychological foundation of our research. To ground our discussion in a specific mathematical context, we use arithmetic learning as a basis for describing how children develop various mathematical strategies. We then turn to issues of children's learning of arithmetical strategies and classroom strategy diversity.

1.1. Intra-individual variability of strategies

As children develop arithmetical competencies, they will use multiple strategies and choose among these strategies adaptively. This process supports their future learning (Coyle, 2001). Learning and development involve an ongoing competition among alternative strategies, with faster and more accurate strategies gradually becoming dominant (Kerkman & Siegler, 1993; Minsky, 1986; Siegler, 1993). Intraindividual variability tends to be greatest during periods of rapid learning (although substantial variability is present in relatively stable periods as well, Siegler, 2006). Intraindividual variability also tends to be cyclical as periods of lesser and greater variability alternate over the course of learning (Siegler, 2006). For instance, a child learning to solve simple addition or subtraction problems by counting all objects versus counting on from one addend may oscillate between the two strategies until the child becomes comfortable exclusively using the more efficient counting on strategy.

One can ask whether possessing a larger number of different strategies as an individual is developmentally advantageous. Siegler (1995) reported that early variability predicts later achievement and posited the moderate experience hypothesis: use of multiple strategies is most likely when people have moderate amounts of experience with the problems being addressed. This leads to an inverted-U for the number, and therefore intra-individual variability, of strategies of individual children.

However, consideration of variability alone may not be sufficient–strategies must be efficient and especially flexible/adaptable (Baroody & Dowker, 2003; Torbeyns, Verschaffel, & Ghesquière, 2005; Verschaffel, Torbeyns, De Smedt, Luwel, & Dooren, 2007). Adaptation of strategies to different contexts is viewed as a positive aspect of mathematics development (Baroody & Dowker, 2003; Lemaire & Siegler, 1995). Children tend to choose strategies adaptively; even if they have had fewer opportunities to learn and thus use less sophisticated strategies, children are as adaptive as children with more sophisticated strategies (Kerkman & Siegler, 1993). Further, when provided opportunities to develop content knowledge, the former children's levels of accuracy, speed, and strategy use were comparable (Siegler, 1993).

1.2. Classroom strategy diversity and children's learning of arithmetic

Although many educators consider strategic development central to learning, their goals for *diverse* strategy use by children differ substantially. Because distinct learning goals have strong effects on strategy development (Clements, Agodini, & Harris, 2013; McNeal, 1995), we consider three categories of approaches to children's learning. Before we define these categories, however, we clarify the relations of our classification structure to the more general framework of discovery learning versus learning from direct instruction to avoid any confusion. Theory and research on the latter framework often use three categories as well: unassisted discovery learning, enhanced discovery learning, and learning from direct or explicit instruction (Alfieri, Brooks, Aldrich, & Tenenbaum, 2010). Our research focuses specifically on classroom strategy diversity with possible implications for pedagogical approaches (although we do not measure teacher instruction here, these will be a proper subset of the larger discovery/explicit instruction debate). However, although our categories relate to this general framework, there is not a direct mapping, as we delineate in this and the final section.

Our first category values as much diversity as possible. Consistent with this, researchers have shown the benefits of teachers' noticing, encouraging, and discussing children's different solution strategies (e.g., Choppin, 2011; Stein, Engle, Smith, & Hughes, 2008). As an example in early arithmetic, some children may solve 3 + 5 with a counting-all procedure, counting out a set of three items, then counting out five more items, and then counting all those starting at "1," reporting "8." Others may count on from the initial number to the total, keeping track of the number added on the fingers, as in "3…5, 6, 7, 8!" Still, others might put up three fingers on one hand and five on the other then recognize the total. This category implies the prediction that more diversity at every time point will correlate with higher achievement at the early time point (concurrently) and at subsequent time points because it supports children's creative thinking and use of strategies that are most meaningful to them (Kam, 2017). Note that in this approach, as well as the other two approaches, such inter-individual diversity does not necessarily imply or depend on intraindividual strategy variability; for example, children may use one "favorite" strategy, but that strategy may differ from child to child.

The second category includes the competing perspective that education is more efficient and mathematically rigorous if children learn accurate definitions and demonstrate a single prescribed, accurate mathematical procedure for a certain type of problem, obviating the need for children to invent or experience multiple strategies (see Carnine, Jitendra, & Silbert, 1997; Clark, Kirschner, & Sweller, 2012; Wu, 2011). Research findings on children's learning is mixed, but there is evidence supporting this approach to children's learning (Carnine et al., 1997; Clark et al., 2012; Gersten, 1985; Heasty, McLaughlin, Williams, & Keenan, 2012). This category implies the prediction that *less* diversity at every time point would correlate with higher achievement at early and subsequent time points.

A third category is a combination of the first two. Here, strategy diversity is an early goal, with subsequent funneling of children's strategic use to more effective strategies (not aimed at diversity per se). For example, in a Japanese approach to teaching first grade arithmetic (Murata & Fuson, 2006), teachers first elicit, value, and discuss child-invented strategies and encourage children to use diverse strategies to solve a variety of problems. However, later in the year, teachers focus on and encourage a particularly effective method (Henry & Brown, 2008; Murata & Fuson, 2006). This approach is consistent with the notion that variability supports early phases of learning, but expertise brings decreasingly variable performance across children (Siegler, 1994). Thus, this category implies the prediction that *early* diversity would correlate with higher achievement at an early time point (for material children were beginning to learn) as well as at subsequent time points, but that at subsequent time points *less* diversity in strategies (e.g., because a class coalesced around fewer, sophisticated strategies) would be positively related concurrently.

This study compared the implications of these three categories. Thus, our research question was: Which of the three categories' prediction is most consistent with the empirical results regarding the relations of strategy diversity in classrooms to children's concurrent and subsequent math achievement? That is, how does classroom-level strategy diversity relate to children's mathematics achievement, both concurrently and at later time points? Such a question has important implications both for research and practice.

1.3. Present study

Although the three categories contribute to the wider debate regarding general categories of unassisted discovery learning, enhanced discovery learning, and learning from direct or explicit instruction (Alfieri et al., 2010), each of the three could be used in tandem with more than one of these general categories, meaning that research on classroom strategy diversity is a theoretically and empirically distinct subfield.

Given that both intraindividual strategy variability and inter-individual strategy diversity are theoretically most important from infancy through the early childhood years (Siegler, 1994), we addressed our research questions in the earliest years of mathematics instruction. Data for this study came from the experimental study of the TRIAD (Technology-enhanced, Research-based, Instruction, Assessment, and professional Development) scale-up model. This model evaluated the implementation of a preschool math program (Clements, Sarama, Spitler, Lange, & Wolfe, 2011; Sarama, Lange, Clements, & Wolfe, 2012) and the first year of implementation of a follow-through intervention at the end of kindergarten (Sarama, Clements, Wolfe, & Spitler, 2012). At the preschool level (called pre-K hereafter to denote the pre-kindergarten year), the two experimental conditions were identical. Evaluations revealed a substantial and significant effect at the end of pre-K (effect size, g = 0.72, Clements et al., 2011). One experimental condition was assigned to experience a follow-through intervention into the kindergarten and first-grade years, in which teachers were taught about the pre-K intervention and ways to build upon it using learning trajectories. Kindergartners in both the follow-through condition (g = 0.38) and non-follow-through condition (g = 0.30) scored statistically significantly higher than children in the control condition, although the effect sizes were about half of that at the end of the pre-K year (Sarama, Clements, et al., 2012). Similarly, first graders in both the follow-through condition scored significantly higher than control children (g = 0.51 for those who received follow through intervention in kindergarten and first grade; g = 0.28 for non-follow through) and follow-through children scored significantly higher than non-follow-through children (g = 0.24, Clements, Sarama, Wolfe, & Spitler, 2013). Analyses used a Rasch score that incorporated two types of scores: correctness for all items and, for those with observable and variable strategies, a strategy-sophistication score. Because the mathematical assessment recorded the specific strategies children used, it allowed us to create a new measure of the diversity of strategies used within each class, at least in an individual assessment context, and relate this measure to the Rasch scores, concurrently and predictively. Most notably, the primary benefit of using extant TRIAD data for the present study is that the implementation of multiple interventions ensured a variety of pedagogical activities and teaching strategies, contributing to our analyses.

2. Method

Data from the TRIAD cluster randomized experiment was used for the current study, which included 42 schools in two cities that were randomly assigned to one of the three conditions.

2.1. Participants

Participants were the 1305 children from the original 42 schools and 106 classrooms in Buffalo, NY, and Boston, MA who had both a pretest and posttest in pre-K (Clements et al., 2011), and the pre-K and kindergarten teachers in those schools. At the pretest of the pre-K year, children ranged in age from 44 to 64 months, with a mean age of 52.06 months (SD = 4.09). 24.5% (n = 319) of children were identified as bilingual, and 51% (n = 664) of them were female. In addition, 2% of children reported their ethnicity as Native American; 53% of children reported their ethnicity as African-American; 4% reported their ethnicity as Asian/Pacific Islander; 22% reported their ethnicity as Hispanic; 19% reported their ethnicity as White non-Hispanic and < 1% other. From the pre-K year to first-grade year, the total attrition rate in this study was 13.74% (n = 179), and the attrition was unrelated to child demographic variables of bilingualism (χ^2 = 0.02, df = 1, p = .89), gender (χ^2 = 0.20, df = 1, p = .65), and race/ethnicity (χ^2 = 6.75, df = 5, p = .24).

2.2. Assessments

The Research-based Elementary Math Assessment (REMA, Clements, Sarama, Wolfe, & Day-Hess, 2008/2019) measures core mathematical abilities of children from age 3–8 years using an individual interview format with standardized administration protocol, videotaping, coding and scoring procedures. Abilities are assessed according to theoretically- and empirically-based developmental progressions (National Research Council, 2007; Sarama & Clements, 2009). Topics in number include verbal counting, object counting, subitizing, number comparison, number sequencing, connection of numerals to quantities, number composition and decomposition, adding and subtracting, and place value. Geometry topics include shape recognition, shape composition and decomposition, congruence, construction of shapes, and spatial imagery, as well as geometric measurement, patterning, and reasoning. The developmental progression of items as well as the fit of individual items has been reported in earlier research (Clements, Sarama, & Liu, 2008). The REMA measures mathematical competence as a latent trait in Item Response Theory (IRT), yielding a score that locates children on a common ability scale with a consistent, justifiable metric (allowing accurate comparisons, even across ages and meaningful comparison of change scores, even when initial scores differ, Wright & Stone, 1979). The 225 items are ordered by Rasch item difficulty; children stop after four consecutive errors on each of the number and geometry sections.

Beyond correctness, the REMA also collects, codes, and scores children's strategies when those are observable and relevant (e.g., processes of perceptual subitizing, or quick recognition of the number of objects in a set, would be neither). Arithmetic problems are a main sources of such strategies. An example item, including strategies that are common to many REMA items, is provided in Fig. 1. As an example of quite different strategies, Fig. 2 presents an item that assesses geometric composition of two-dimensional geometric shapes. These strategies are recoded into three levels of sophistication for submission to the Rasch model. Until this study, no other analyses were made of the diversity of the strategies.

Training sessions for REMA assessors included orientation, demonstration, and practice, with a focus on standardized delivery. Subsequent individual practice sessions were taped and critiqued, with 98–100% error-free delivery required for certification. All assessment sessions were videotaped, and each item coded by a trained coder for

Have *no* objects available. *Note.* Allow, but do not mention or encourage, use of fingers.

How much is 6 + 4?
¿Cuánto es seis (6) más cuatro (4)?

Correct: 10

Code 59A:
0 = incorrect
1 = correct (10)
9 = no response

Code 59B: What did the child do?
1 = counted all objects (fingers or other impromptu items), specifically counted out 6 (or 4), then 4 (or 6), then counted all starting at 1; concrete modeler (e.g., shows 6 fingers, then 4 more, says 10); includes transitional strategies such as adding on, counting 6 fingers and then putting up 4 more fingers on the other hand while continuing the count
2 = counted on from 6 or 4 (e.g., said 6, then counted out or put up 4, counting, 7, 8, 9, 10) *or* kept track while counting on (e.g., said 6, then counted, 7, 8, 9, 10 while putting up 4 fingers)
3 = verbalized a derived combination (6 + 2 = 8, plus 2 more is 10)
4 = verbalized a combination (6 + 4 = 10)
7 = other strategy (describe on Assessment Record)
8 = strategy not observed
9 = NA

Fig. 1. Example arithmetic item with strategies.

correctness and for solution strategy; 10% of the assessments were double-coded. Both assessors and coders were blind to the group membership of the children. Continuous coder calibration by an expert coder (one tape per coder per week) militated against drift. Calibration feedback was sent to coders, alerting them to any variance from coding protocols. Previous analysis of the assessment data showed that its reliability ranged from 0.93 to 0.94 on the total test scores (Clements et al., 2008); the reliability was 0.92 with the present population. In addition, the REMA had a correlation of 0.86 with a different measure of preschool mathematics achievement (Clements et al., 2008), the Child Math Assessment: Preschool Battery (Klein, Starkey, & Wakeley, 2000), and a correlation of 0.74 with the Woodcock-Johnson Applied Problems subscale for a pre-K specific subset of 19 items (Weiland et al., 2012).

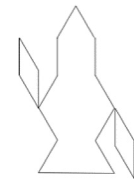
2.3. Intervention

We briefly describe the intervention used in the original experiment to elucidate the level of understanding early childhood teachers in our sample possessed regarding mathematics instruction. The TRIAD intervention helps schools implement the *Building Blocks* pre-K mathematics curriculum (Clements & Sarama, 2007/2013) that was designed using a comprehensive Curriculum Research Framework (Clements, 2007) to address numeric/quantitative and geometric/spatial ideas and skills. Woven throughout the *Building Blocks* curriculum are mathematical processes. That is, *Building Blocks* is structured on learning trajectories that were designed to develop teachers' content knowledge by explicating the mathematical concepts, principles, and processes involved in each level and the relationships across levels and topics. As an arithmetical example, children may first "count all" as in solving a story problem involving 3 + 2 by counting out three chips, then counting out two, then counting all five starting at 1. They then extend their counting skills to larger numbers, enabling them to solve problems such as 5 + 3. At a higher level in the developmental progression, they understand that starting a counting sequence *at* the first number can "stand in for" counting out the first addend–the number word substitutes for the counting acts from 1 to 5. That is, saying "fiiiive…" is a way of reifying that quantity. Children can then count on from that number, counting on through second addend ("fiiiive…, *six*, *seven*, *eight*. Eight!"). They may use a "rhythm of three" ("Doo – Day – Doo") to create the sequence "six, seven, eight." At the following level of the developmental progression, children use anticipatory thinking to keep track of those counts (especially when a rhythmic pattern is difficult, such as adding 5 + 7). For example, starting at 5 and counting seven counting words, while putting one finger up for each count until a finger pattern of 5 (five fingers on one hand, two on the other) is reached (or putting up seven fingers to begin, and lowering one finger for each count). Later, children learning de/composition strategies, such as 5 + 7 = 5 + 5 + 2 = 10 + 2 = 12 and so forth.

Such learning trajectories were intended to develop teachers' knowledge of children's developmental progressions in learning that content. They were designed to develop teachers' knowledge of the instructional activities created to teach children the content and processes defining the level of thinking in those progressions and to inform teachers of the rationale for the instructional design of each (e.g., why the curriculum teaches children to start at numbers other than one and keep counting and why it shows and then hides chips representing the first addend to encourage counting on). The learning trajectories assist curriculum enactment with fidelity in that they connect the developmental progressions to the instructional tasks, providing multiple guidelines or sources of stability in teachers' instantiation of the instructional activities. They were also designed to motivate and support the use of formative assessment. The TRIAD scale-up model included extensive professional development, both training and coaching (for details see Clements et al., 2011).

Kindergarten and first-grade teachers in schools assigned to TRIAD's

Give the child the set of pattern blocks, randomly mixed in front of them, and the picture of a puzzle (right). Say: "**Use pattern blocks to fill this puzzle. Put them together with full sides touching.**"



Code 1A (*Very small gaps or misalignments that can be attributed to fine motor limitations are acceptable*)

0 = incorrect (placed no shapes *or* placed shapes but not one "fit" the puzzle form, where *fit = at least one side aligned, with no "hangover" outside the puzzle.)*
1 = "partially correct" (one or more shapes "fit" but there were one or more gaps or "hangovers")
2 = correct (completed puzzle accurately; no gaps or "hangovers")
NR = no response

Code 1B *For all but 1-2 of the shapes,*

0 = selection of shapes not focused on completing puzzle (e.g., selects all red trapezoids)
1 = was hesitant or not systematic (e.g., used cycles of trial and error)
2 = completed the puzzle correctly, systematically, but may be "halting"
3 = completed the puzzle correctly, immediately, and confidently
9 = NR (no response)

Code 1C *For all but 1-2 of the shapes,*

0 = selection of shapes not focused on completing puzzle (e.g., selects all red trapezoids)
1 = turned shapes after placing on puzzle in an attempt to get them to fit
2 = turned shapes into correct orientation prior to placing them on the puzzle
9 = NR

Code 1D *For all but 1-2 of the shapes,*

0 = selection of shapes not focused on completing puzzle (e.g., selects all red trapezoids)
1 = tried out shapes by picking them seemingly at random, then putting them back if they did not look right, so seemingly trial and error
2 = appeared to search for "just the right shape" that they "know will fit" and then finding and placing it.
9 = NR

Figure 2. Example geometry item with strategies.

Fig. 2. Example geometry item with strategies.

follow-through condition were taught about the pre-K intervention and ways to build upon it. That is, they were shown the mathematics many of their entering children had learned. Teachers were also taught about the learning trajectories that extended through the kindergarten curriculum, including the developmental progressions and how to modify their extant curricula to more closely match the levels of thinking of their children; however, they were not provided with any specific activities (for details see Sarama, Clements, et al., 2012). Although discussion of strategies was encouraged in the TRIAD intervention, there was no focus on diversity of strategies per se at any grade, so none of the treatment conditions were a priori aligned with the three categories. (Further, the impact of intervention on diversity scores across four time-points is small; at T_1 and T_3 they are statistically significant, but not at T_2 and T_4. Most important, the effect sizes were small; partial η^2 ranges from 0.01 to 0.07). The control teachers taught a curriculum that, like the treatment conditions, was connected to standards and thus addressed the same topics, in approximately the same proportion.

Of course, teachers in all three treatment conditions taught mathematics differently, with variance in their evocation, discussion, and support of different mathematical strategies (as indicated by teacher observations, Clements et al., 2011; Clements et al., 2013; Sarama, Clements, et al., 2012).

2.4. Teacher training and classroom observations

Pre-K teachers in schools randomly selected to receive the intervention completed seven days of professional development (PD) the year before data collection, during which time they implemented the Building Blocks curriculum in a non-stressful context. They received an additional five days of professional development during the following year. They also received coaching about two times per month for both years. The kindergarten teachers received seven days of PD that started in the fall when they received the cohort from the preschool intervention. Assessors administered the REMA to all children in the fall of their pre-K year before the intervention started, at the end of their pre-K year, and at the end of their kindergarten year.

2.5. Quantification of strategic diversity

Given that literature currently focuses on intra-individual strategy variability, addressing our research questions meant facing the methodological challenge of quantifying inter-individual strategy diversity displayed by children within a particular classroom or school. Recall that the REMA was designed to be flexibly administered to children of varying levels of mathematical competence and it features previously validated (Clements et al., 2008, 2019) start- and stop-rules that allow the test-administrator to determine which REMA items a child will attempt, given their ability level. For practical considerations in the creation of a scoring algorithm for the quantification of strategic diversity, this aspect of the REMA meant that participants each attempted different items, and different numbers of items at that. In addition, the codes used by test administrators to indicate which strategy a child utilized were nominal in nature and reflected a unique way to solve that math problem (i.e., they were not ordinal, although in later stages of analyses they were recoded into three levels of sophistication for Rasch scoring.). Further, the REMA items also featured different total numbers of possible strategies, such that a single set of nominally coded strategic processes was impossible to define across all the items.

Although item response theory (IRT) methods for estimating a latent psychological attribute from nominal data exist (Darrell Bock, 1972), these methods are more adept at estimating latent attributes in relatively large sample sizes both in terms of participants and the number of items administered. Nominal IRT approaches also produce more readily interpretable scores if all participants are administered the same items and those items feature the same nominal categories (Nering & Ostini, 2011), which are not characteristics of our REMA data. In addition, some items on the REMA have 10 or more possible strategic process codes, which would be more nominal categories than are typically recommended for use in a nominal IRT model. Moreover, the focus of this investigation is not on the latent attribute of strategic ability per se, but on the inter-individual variance in the deployment of strategies within a salient group of children (e.g., a classroom). Although nominal IRT models estimate variances for the latent attributes being measured, that latent variance would represent the degree to which children varied along the latent continuum, and would not necessarily indicate strategic diversity within the classroom in the way that is intended in this investigation. Therefore, we conceptualized and developed a novel procedure for quantifying strategic diversity within a classroom based on REMA nominal strategy codes.

To enact this method, the existing REMA strategy codes for every item (j) were dis-aggregated into a number of new dichotomously coded variables, where zero indicated the absence of that strategy, and 1 indicated the use of that strategy. In this way, the number of possible strategy codes that were available for any given item dictated the number of newly created dichotomously coded strategy variables (k) for that REMA item. Then, following a common distributional assumption in psychometrics (e.g., Bauer & Hussong, 2009), each of these new dichotomously scored strategy variables were assumed to follow a Bernoulli distribution, a special-case of the binomial distribution for a single variable in which the variance of that dichotomously coded variable is represented by Eq. (1), where p is the mean of the dichotomously coded variable (i.e., the proportion of 1′s coded for that variable in the dataset). Because this investigation was focused on the creation of diversity scores that represented the variety of strategies employed by children within each particular classroom in the dataset, these strategy variances were calculated individually for every dichotomously coded strategy variable (k), which had been dis-aggregated from every REMA item (j) within each individual classroom (c) in the dataset following Eq. (1), below:

$$\sigma^2_{jkc} = p_{jkc} \times (1 - p_{jkc}) \tag{1}$$

After calculating these variances following Eq. (1), they were summed within every classroom (c) to create a quantity that represented the total variance on the strategy codes within that classroom. Then, to correct for the fact that classrooms differed in the number of items (N) that children attempted as well as the number of strategies that were available for each of those items, the summed variances were divided by the median number of dichotomously scored strategy variables that were available on the REMA to the children within that classroom. Finally, the resulting decimal was multiplied by 100 for scale. In effect, this strategy diversity score is the degree to which children within a particular classroom varied inter-individually on their strategy usage on the items they were exposed to. The calculation of this *diversity score* is presented in Eq. (2).

$$DS_c = \left(\frac{\sum_{i=1}^{N} \sigma^2_{jkc}}{P_{50}(N_c)} \right) \times 100 \tag{2}$$

These strategy diversity scores were saved for every classroom in the analytic dataset and analyzed as described in the following Results section. It should be noted that the diversity scores have the common property of being influenced by the number of individual children within a given classroom. For instance, hypothetically, if only one child were present in a classroom, the diversity score would necessarily be zero because inter-individual variance is impossible in that case. Given the focus of this investigation on inter-individual diversity in strategic processing, such a property of the diversity scores was not considered problematic; however, further analysis was limited to classrooms with at least five study children to provide meaningful inferences about strategy diversity.

3. Results

Correlations and multi-level linear regression analyses were conducted to examine the three predictions regarding relations of strategy diversity in classrooms to children's concurrent and subsequent math achievement. To address our research questions, we conducted an analysis of the resulting classroom strategy diversity scores that unfolded in three general stages. First, we examined changes in strategy diversity across time points (i.e. pre-K through first grade) through a comparison of the average diversity scores across classrooms within each time-point. Then, we tested bivariate correlations among strategy diversity and standardized mathematics achievement across the four time-points included in this study. Third, we conducted two-level linear models with random intercepts to examine the predictive power of classroom diversity scores in predicting mathematics achievement at each time-point, while also adjusting for the classroom-nested structure of these educational data. Furthermore, multi-level models incorporating child background characteristics as level-one covariates (i.e., baseline mathematics achievement, age, socio-economic status,

Table 1
Descriptive statistics and bivariate correlations.

Variable		Mean	SD	Bivariate Correlations							
				Math Achievement				Strategy Diversity			
				T1	*T2*	*T3*	*T4*	*T1*	*T2*	*T3*	*T4*
Math Achievement	*T1*	−3.22	0.82	1.00							
	T2	−1.97	0.71	0.57**	1.00						
	T3	−1.09	0.68	0.55**	0.76**	1.00					
	T4	−0.07	0.70	0.54**	0.73**	0.81**	1.00				
Strategy Diversity	*T1*	11.13	8.31	0.24**	0.26**	0.27**	0.26**	1.00			
	T2	15.02	5.16	0.14**	0.11**	0.17**	0.17**	0.27**	1.00		
	T3	10.87	3.67	−0.11**	−0.12**	−0.16**	−0.19**	−0.07*	0.01	1.00	
	T4	9.33	2.51	−0.09*	−0.03	−0.06	−0.08*	0.06	0.03	0.18**	1.00

Note. T1 = pretest of pre-K year; T2 = posttest of pre-K year; T3 = posttest of Kindergarten year; T4 = posttest of 1st grade year.
** $p < .01$.
* $p < .05$.

intervention versus control condition, and bilingualism) were performed to fully model the pattern of influence of classroom-level strategic diversity scores on math achievement. Results of each stage of this investigation follow.

3.1. Score description

Table 1 presents means and standard deviations of math achievement and diversity scores at four-time points: pre- and posttest of the pre-K year (T_1 and T_2) and posttests at the end of the kindergarten and first-grade years (T_3 and T_4). As can be seen, children's average standardized math scores gradually increased across time (*M* range = −3.22 to −0.07, *SD* range = 0.68–0.82) while the average classroom strategy diversity scores, as well as their level of variation (i.e., standard deviation), decreased from the end of the pre-K year to the first-grade year (*M* range = 9.33–11.13, *SD* range = 2.51–8.31). The decreases in diversity scores were statistically significant, [F (2, 1458) = 410.79, $p < 0.001$, partial $\eta^2 = 0.36$]. In other words, strategy use within classrooms was significantly more diverse in the earlier learning phases (T_1 and T_2) than the later learning phases (T_3 and T_4), and, meanwhile, the level of strategy diversity across different classrooms became less heterogenous (i.e., SDs decreased).

3.2. Bivariate correlations

3.2.1. Concurrent time point analysis

The bivariate correlations (presented in Table 1) revealed positive and significant correlations between classroom strategy diversity and concurrent achievement at pretest ($r = 0.24$, $p < .01$) and posttest ($r = 0.11$, $p < .01$) in the pre-K year. However, at the later time points (i.e., the end of kindergarten and end of first grade) diversity in strategies turned to be negatively correlated with the concurrent math achievement scores ($r = -0.16$, $p < .01$ for kindergarten; $r = -0.08$, $p < .05$ for first grade).

3.2.2. Subsequent time point analysis

Furthermore, diversity scores at the pre-K year positively and significantly correlated with math achievement at subsequent time points (r ranging from 0.17 to 0.27, $p < .01$) while strategy diversity of the kindergarten year negatively correlated with achievement at 1st grade ($r = -0.19$, $p < .01$).

3.3. Multi-level linear regression analyses

The intraclass correlations (ICC, Raudenbush & Bryk, 2002) in the predictions of child-level math achievement at the three time-points (T_2, T_3, and T_4) were 19.29%, 26.13%, and 35.18%, respectively, which indicated that a non-trivial amount of the variation in math achievement occurred between classrooms. Therefore, multi-level regression analysis was considered appropriate to account for the classroom-nested structure of these data.

3.3.1. Means-as-outcomes models

To initially investigate how diversity scores can contribute to predicting math achievement at the classroom level, means-as-outcomes models that only contained the classroom-level predictor (i.e., classroom strategic diversity scores) were performed. See Table 2 for the specific coefficients discussed in this section. After controlling for the classroom diversity scores, the clustering effects were reduced (conditional ICC_{T2} = 13.26%, conditional ICC_{T3} = 14.29%, and conditional ICC_{T4} = 13.26%), which indicated that less variation of math achievement occurred between classrooms at each time-point, after accounting for classroom level strategic diversity.

In the analysis predicting math achievement at the end of the pre-K year, the proportion of variance explained (PVE) by the model with diversity scores at T_1 and T_2 was equal to 35.93%. That is, about 36% of the between-classroom variance in math achievement was accounted for by the combination of diversity scores at two time-points. Specifically, more classroom strategy diversity at the pretest of pre-K predicted higher achievement scores at the pre-K posttest ($B = 0.021$, $SE = 0.004$, $z = 5.57$, $p < .001$), and the concurrent diversity score of the pre-K posttest was not a significant predictor.

Then, in the multi-level regression analysis predicting math achievement at the end of the kindergarten year, a larger amount of the true between-classroom variance (PVE = 55.57%) in math achievement was accounted by diversity scores at T_1, T_2, and T_3. The three diversity scores were all significant predictors of math achievement at the end of the kindergarten year ($B_{T1} = 0.015$, $SE = 0.003$, $z = 4.84$, $p < .001$; $B_{T2} = 0.012$, $SE = 0.005$, $z = 2.24$, $p = .025$; $B_{T3} = -0.027$, $SE = 0.008$, $z = -3.34$, $p = .001$). However, they yielded different directions in that prediction. Specifically, children from classrooms with more strategy diversity in the earlier learning phases (pre- and posttest of the pre-K year) demonstrated higher math achievement at the end of kindergarten, but children in classrooms with higher diversity at the end of kindergarten demonstrated lower concurrent achievement.

Lastly, in the analysis predicting math achievement at the end of the first-grade year, an increasingly larger amount of the between-classroom variance (PVE = 59.78%) in math achievement was accounted for by classroom diversity scores at T_1, T_2, T_3, and T_4. The diversity scores at the pre-K and kindergarten year all predicted first-grade math achievement ($B_{T1} = 0.015$, $SE = 0.004$, $z = 4.12$, $p < .001$; $B_{T2} = 0.013$, $SE = 0.006$, $z = 2.28$, $p = .023$; $B_{T3} = -0.024$, $SE = 0.007$, $z = -3.35$, $p = .001$), while the concurrent diversity score

Table 2
Summary of multi-level linear regression predicting early mathematics (Means-as-Outcomes Models, Level 2 Predictors only).

Parameters	*Estimates*	*SE*	*z*	*p*	PVE	ρ ($\hat{\rho}$)
Predicting Math Achievement at the End of Pre-K Year						
Fixed Effects						
Diversity Score T1	0.021	0.004	5.570	< 0.001***		
Diversity Score T2	0.007	0.006	1.130	0.259		
Random Effects (variance components)						
Intercept, μ_{0j}	0.063	0.014	4.597	< 0.001***	35.93%	19.29%
Residual, r_{ij}	0.409	0.017	24.366	< 0.001***		(13.26%)
Predicting Math Achievement at the End of Kindergarten Year						
Fixed Effects						
Diversity Score T1	0.015	0.003	4.840	< 0.001***		
Diversity Score T2	0.012	0.005	2.240	0.025*		
Diversity Score T3	−0.027	0.008	−3.340	0.001**		
Random Effects (variance components)						
Intercept, μ_{0j}	0.055	0.013	4.143	< 0.001***	55.57%	26.13%
Residual, r_{ij}	0.332	0.016	20.402	< 0.001***		(14.29%)
Predicting Math Achievement at the End of First-Grade Year						
Fixed Effects						
Diversity Score T1	0.015	0.004	4.120	< 0.001***		
Diversity Score T2	0.013	0.006	2.280	0.023*		
Diversity Score T3	−0.024	0.007	−3.350	0.001**		
Diversity Score T4	−0.017	0.015	−1.110	0.266		
Random Effects (variance components)						
Intercept, μ_{0j}	0.074	0.018	4.015	< 0.001***	59.78%	35.81%
Residual, r_{ij}	0.325	0.019	17.381	< 0.001***		(18.56%)

Note. T1 = pretest of pre-K year; T2 = posttest of pre-K year; T3 = posttest of kindergarten year; T4 = posttest of first-grade year, PVE = portion of variance explained, ρ = *intraclass correlation,* $\hat{\rho}$ = conditional intraclass correlation.
*** $p < .001$.
** $p < .01$.
* $p < .05$.

at T_4 was not a significant predictor (B_{T4} = .−017, SE = 0.015, $z = -1.11$, $p = .266$). In terms of the directions of these coefficients, more classroom strategy diversity in the earlier learning phases (T_1 and T_2) and less during the later learning phases (T_3 and T_4) appears to support children's math achievement, given the results of this classroom level model.

3.3.2. *Full multi-level models with random intercepts*

To further verify the observed patterns of mathematics achievement prediction, baseline math scores, age, socio-economic status, intervention versus control condition, and bilingualism, were included as child-level covariates in the two-level models. The mean-centering of predictor variables within multi-level models in educational research has been a much-discussed topic within the methodological literature (for technical treatments of this issue see Plewis, 1989; Raudenbush, 1989). However, there is currently not a clear rule of thumb for this modeling decision (Paccagnella, 2006). In this study, we did not impose centering strategies to the continuous predictors at level-one for a number of reasons. For example, interpreting level-one predictors is not a primary goal of the present study, and there was no strong collinearity among all predictors, making mean-centering extraneous. Table 3 presents the specific coefficients discussed in this section.

Compared to the previous models with classroom-level predictors only, the significance and strength of the coefficients associated with the diversity scores was weakened after accounting for all of these child-level control variables in predicting math achievement of children in kindergarten and first-grade years. The child-level covariates altered the pattern of significance of the diversity scores in predicting pre-K math achievement (see Table 3 for specific coefficients). Nevertheless, the directions of diversity score parameters from the kindergarten and first grade regression models remained as in the previous model, which provided further evidence for the observed pattern that more strategy diversity in the earlier years (T_1 and T_2) and less in the later years (T_3 and/or T_4) within math classrooms could help children achieve better math learning outcomes, even after controlling for a variety of child-level covariates.

4. Discussion and implications

Most educators agree that the strategies children use to solve mathematical problems are an important component of their learning (Biddlecomb & Carr, 2011; Carpenter et al., 1998; Carr et al., 1999; Fennema et al., 1996; Pang & Kim, 2018; Sherin & Fuson, 2005; Wansart, 1990). Pedagogically, instruction focused on arithmetic strategies can be effective for all children (Franke et al., 2007; Jacobs & Empson, 2016) including, and perhaps especially, for those with special needs (e.g., Fuchs et al., 2010; Naglieri & Johnson, 2000; Powell & Fuchs, 2015; Rockwell et al., 2011). However, there is little research that investigates whether classrooms that evince a broader diversity of strategies engender higher mathematics achievement. We begin by discussing the results in terms of the predictions of the three categories of approaches to children's learning of strategies, especially diverse strategies. Although we did not measure teachers' instruction, the results have important implications because there are quite different pedagogical approaches for teaching strategies to children that have different goals and approaches have a strong effect on strategy development (Clements, Agodini, & Harris, 2013; McNeal, 1995). Therefore, we discuss the pedagogies used in each of the three categories and draw implications for them from our findings.

4.1. *Classroom strategy diversity predicting children's learning*

We examined patterns between class-level diversity and children's achievement in light of three possible categories. The first category values strategy diversity at every time point of mathematical learning; the second category suggests that children should use a single, preferred

Table 3
Summary of multi-level linear regression analysis (full model) early mathematics.

Parameters	*Estimates*	*SE*	*z*	*p*
Predicting Math Achievement at the End of Pre-K Year				
Level 1 Predictors				
Baseline	0.436	0.022	19.890	< 0.001***
Age	0.026	0.004	5.840	< 0.001***
Bilingual	0.100	0.042	2.390	0.017*
FRL	−0.155	0.045	−3.430	0.001**
Intervention	0.424	0.052	8.110	< 0.001***
Level 2 Predictors				
Diversity Score T1	0.004	0.003	1.250	0.212
Diversity Score T2	−0.003	0.005	−0.680	0.497
Predicting Math Achievement at the End of Kindergarten Year				
Level 1 Predictors				
Baseline	0.686	0.024	28.360	< 0.001***
Age	0.009	0.004	2.200	0.028*
Bilingual	0.139	0.037	3.770	< 0.001***
FRL	−0.127	0.042	−3.050	0.002**
Intervention	−0.114	0.045	−2.500	0.012*
Level 2 Predictors				
Diversity Score T1	0.005	0.002	1.990	0.046*
Diversity Score T2	0.006	0.004	1.720	0.086
Diversity Score T3	−0.008	0.005	−1.460	0.144
Predicting Math Achievement at the End of First-Grade Year				
Level 1 Predictors				
Baseline	0.770	0.027	28.150	< 0.001***
Age	−0.003	0.004	−0.780	0.437
Bilingual	0.008	0.040	0.190	0.848
FRL	−0.217	0.044	−4.900	< 0.001***
Intervention	−0.023	0.042	−0.540	0.586
Level 2 Predictors				
Diversity Score T1	0.003	0.002	1.070	0.286
Diversity Score T2	0.004	0.004	1.250	0.210
Diversity Score T3	−0.008	0.005	−1.750	0.080
Diversity Score T4	−0.016	0.007	−2.110	0.035*

Note. T1 = pretest of pre-K year; T2 = posttest of pre-K year; T3 = posttest of kindergarten year; T4 = posttest of first-grade year, FRL = Free/Reduced price lunch status.

*** $p < .001$.

** $p < .01$.

* $p < .05$. Bilingual, FRL, and whether received TRIAD Intervention were coded as 1(Yes) and 0 (No).

strategy and thus evince less diversity at the classroom-level; the third category indicates that children will use a diverse array of strategies when first learning material and then coalesce around a smaller set of strategies as they master material—and thus, early diversity would be related to mathematical achievement at subsequent time points. We used multi-level regression analyses to (1) investigate how diversity scores can contribute to predicting math achievement at the classroom level and to (2) confirm the patterns of mathematics achievement prediction after controlling for child-level covariates including baseline math scores, age, socio-economic status, intervention versus control condition, and bilingualism using two-level models. Both sets of models reveal similar patterns between class-level diversity scores and children's achievement.

We found that children's standardized mathematics scores increased through the four time points, but the average classroom strategy diversity scores, as well as their level of variation, increased in the preschool year, and then decreased at the two subsequent time points.

In the first set of means-as-outcomes models, pre-K fall strategy diversity was positively and significantly related to spring achievement in the pre-K and kindergarten years. Additionally, kindergarten diversity scores were positively and significantly related to first-grade spring achievement. Taken together, these findings imply that children from classrooms with higher levels of diverse child-generated strategies at early time points have higher math achievement scores.

The same set of means-as-outcomes models also indicated that diversity was either unrelated or negatively and significantly related to concurrent achievement (not including the pre-K pretest). For instance, in the pre-K year, spring (posttest) diversity was positively yet nonsignificantly related to spring pre-K math achievement. The same pattern held for spring first grade diversity and spring first grade achievement. In kindergarten, spring diversity was significantly and negatively related to spring achievement. In comparison to the previously described findings between early diversity and later achievement, these findings suggest that concurrent diversity does appear to have the same benefits for concurrent achievement.

In the second set of multi-level models, we found that after controlling for baseline score, age, intervention, socio-economic status, and bilingual status, greater classroom strategy diversity at the pretest of pre-K was positively but not significantly related to higher achievement scores at the pre-K posttest. Additionally, the concurrent diversity score of the pre-K posttest was not a significant predictor. Higher diversity in the fall of pre-K predicted higher achievement in the spring of kindergarten, after controlling for the same set of covariates, but concurrent spring diversity in kindergarten was not related to spring achievement. Counter to our earlier findings, only spring first-grade diversity predicted spring first-grade achievement. Recall that these results accounted for a substantial portion of the variance, even after controlling for a variety of child-level covariates.

These analyses, therefore, provide support most consistent with the approach of the third category. That is, *early* diversity predicted higher achievement for material that children were just beginning to learn as well as their subsequent learning. However, in the later phases of learning, *less* diversity in strategies was positively related concurrently, perhaps because the children in a class coalesced around fewer, more sophisticated strategies.

4.2. Teaching arithmetical strategies and classroom strategy diversity

Each of the three categories not only make unique predictions regarding the relationship of strategy diversity in classrooms to children's concurrent and subsequent achievement but have very different implications for curriculum and teaching. These are important, given the variety of messages teachers receive regarding the type of instruction (i.e., ranging from direct instruction to unassisted discovery) that is best for children's learning. To explicate these pedagogical implications, we use a gardening metaphor that has the goal of creating healthy and beautiful trees and shrubs.

The first category aligns with *unchecked development*, which posits the benefits of providing nutrients and sunshine and otherwise allowing the plants to grow as they will naturally. In education, this pedagogical approach values as much strategy diversity as possible and favors a consistent evocation and sharing of a variety of child-generated strategies and alternative strategies (e.g., Carpenter, Fennema, Franke, Levi, & Empson, 2014; Choppin, 2011; Stein et al., 2008). A goal of such constructivist-based approaches is for children to use and share such different strategies (e.g., Carpenter et al., 2014), thus encouraging classroom strategy diversity, with beneficial effects on learning, presumably based on increases in conceptual understanding (e.g., observing that different strategies yield the same mathematical result) and access to different, often more effective and efficient strategies (Carpenter et al., 1998; Gersten, Beckmann, Clarke, Foegen, Marsh, Star, & Witzel, 2009; Peterson, Carpenter, & Fennema, 1989). This pedagogical approach was advised by Paulos (1991): Stress a few basic principles and leave most of the details to the student (from Baroody, Lai, & Mix, 2006; 1991). This approach may appear aligned with unassisted discovery learning, but it could easily be used in enhanced discovery approaches using the enhancements of generation and elicited explanations (Alfieri et al., 2010).

The second category aligns with *shearing,* or consistently cutting only the top bits of growth to shape the plants into formal hedges for aesthetic or privacy purposes. In education, this competing pedagogy promotes direct instruction as more efficient and mathematically rigorous (see Carnine et al., 1997; Clark et al., 2012; Wu, 2011). In a similar vein, many educators believe that one must teach grade-level goals directly, in contrast to an approach that considers learning trajectories (e.g., see Clements & Sarama, 2014; Sarama & Clements, 2009) in which instruction is based on levels of thinking that most children develop as they build up to competence in such grade-level goals (e.g., Confrey, 2019). The research on such direct teaching and enforcement of specific efficient strategies is contradictory. There is a good body of evidence supporting this approach (Carnine et al., 1997; Clark et al., 2012; Gersten, 1985; Heasty et al., 2012); however, others indicate that teaching formal strategies such as arithmetical algorithms, even with manipulatives, leaves some children even into fourth grade with limited conceptual structures and alternative strategies (Biddlecomb & Carr, 2011).

Supporting a more moderate implementation of this second category, studies suggest that explicitly teaching specific strategies can have a positive impact on children's numerical and arithmetical knowledge (Clarke, Doabler, Nelson, & Shanley, 2014), such as specific teaching of the counting on procedure (Saxton & Cakir, 2006) or teaching a specific sequence of strategies (e.g., Thornton, 1979). These approaches are well aligned with general explicit instruction, although their goals differ as far as teaching one or multiple strategies (Alfieri et al., 2010).

The third category aligns with *pruning*, a synthesis of these two. That is, the goal in gardening is to support natural growth but also to selectively trim branches that are not working and develop an increasingly healthy plant with an aesthetic structure. In education, this would favor an early encouragement and support of child-generated strategies, and only considerably later selectively focusing support on new or strong branches, thus allowing (children's) abandonment of branches that are not growing or sustaining as well. Thus, pruning combines pedagogical approaches that encourage strategy diversity with subsequent support for children to construct more effective strategies (not aimed at diversity per se). Elaborating on the Japanese approach to teaching first-grade arithmetic mentioned previously, teachers first encourage child- invented strategies then focus on and support a particularly effective method later in the year (Henry & Brown, 2008; Murata & Fuson, 2006), in particular the break-apart to make ten strategy (BAMT, such as 9 + 6—what is needed to add to 9 to make 10, subtract 1 from the 6, then add 10 and 5). The BAMT strategy is carefully extended to more challenging numbers and reintroduced in additional contexts. Thus, the classrooms start with a diversity of strategies, but then narrow their focus to shared, effective strategies. This approach does not align with discovery, enhanced discovery, or explicit instruction (Alfieri et al., 2010), but combines these differently at separate phases of the learning process.

In most cases, results from this study aligned more with the pruning approach than with the unchecked development or the shearing approaches. That is, across both sets of models, the earliest phases of learning classroom strategy diversity tended to be positively related to concurrent achievement (contrary to the shearing approach). However, at subsequent time points strategy diversity was usually negatively related or unrelated to concurrent achievement (contrary to the unchecked development approach). Classroom strategy diversity was predictive of achievement from one to three time periods in the future (again contrary to the shearing approach). These relationships were consistent with the implications of pruning approach with one exception in the multi-level linear regression models: we expected kindergarten classroom diversity to positively and significantly predict first-grade achievement. It may be that the range of problems that most children attempted was too narrow and thus restricted the achievement measure. If true, future research examining a wider range of more difficult (e.g., multiplicative) arithmetic problems may reveal different patterns. We note that in the means-as-outcomes models, patterns of findings conformed more precisely to the pruning approach.

It is easy for a lay person to see the results of pruning and shearing as indistinguishable in both gardening and education. However, the processes and the effects are distinct. Shearing is a more rigid intervention, including cutting the leader, selecting a new terminal bud, setting the taper of the tree, trimming side branches, and removing growth seen as problematic because it does not fit a pre-determined shape. It can leave mutilated branches and restrict new growth to the outside of the plants, which results in dead, unhealthy, leafless interiors. Educationally, shearing may help children use mathematically accurate procedures, but curiosity, creativity, and individual development may be curtailed (Baroody & Dowker, 2003; Clements & Sarama, 2014; Kamii & Dominick, 1998; Steffe, 1994; Torbeyns et al., 2005). Pruning, on the other hand, adheres to the plant's natural growth habit and selectively abandons branches anywhere in the shrub where growth is stalled, thus opening up the center of the plant, allowing air and light to maintain healthy growth. In education, then, pruning supports children in creatively inventing their own procedures and then later letting go of unsuccessful and less sophisticated strategies to develop those that are both consistent with children's natural development and stronger mathematically. This could be done via direct instruction but may be more positively with enhanced-discovery instruction featuring carefully selected and scaffolded arithmetical problems, opportunities for children to explain their ideas, and the provision of timely feedback (Alfieri et al., 2010). In both gardening and education, pruning is a more difficult process than shearing.

4.3. Caveats and future research

This is an exploratory study and there are several limitations. First, while the data justify claims about the amount of strategy diversity evident in children's responses, we did not specifically measure teacher instruction of various mathematical strategies[1]. Future research can examine the links between strategy instruction, class-level strategy diversity, and children's mathematical achievement over time. Additionally, although the broader study was a cluster randomized trial, classroom strategy diversity was not randomly assigned, and thus results are only correlational. Further, diversity was examined in children's responses to an assessment. Future research, therefore, could examine children's strategies during instruction, as well as randomly assigning classrooms to interventions embodying the three pedagogical approaches identified here. Such research might also investigate generalizability to other-aged children and thus other topics and levels of mathematical content. Of course, not all instruments collect data on strategies; the present results suggest this may be a useful assessment practice for future research.

4.4. Summary

Keeping these caveats in mind, the findings have important implications for research and practice. They do not support one goal for strategy diversity or a single pedagogical approach. Indeed, they support the pruning approach that synergistically combines the other two approaches and allows multiple general pedagogical stances. Future research is needed, on strategy diversity per se as in the present study as well as via experimental studies with treatments that align to the three pedagogical approaches, unchecked development, shearing, and pruning. We also caution that some children will likely do quite well in an unchecked development environment and perhaps others with shearing approaches. Nevertheless, the evidence of this study suggests

[1] Including, as one reviewer pointed out, variables such as other aspects of teachers' behavior that may be related but not dealing with strategies explicitly, such as autonomy-supportive versus controlling behavior.

that the combination of approaches—children's invention in a guided discovery setting with teachers who encourage diverse strategy use for a considerable period, followed by more focused support on mathematically sophisticated and effective strategies–may be superior for most young children.

Declaration of Competing Interest

The authors declare that they have no conflicts of interest.

References

Alexander, P. A., Grossnickle, E. M., Dumas, D., & Hattan, C. (2018). A retrospective and prospective examination of cognitive strategies and academic development: Where have we come in twenty-five years? In A. O'Donnell (Ed.). *Oxford Handbook of Educational Psychology*. Oxford, UK: Oxford University Press.

Alexander, P. A., & Judy, J. E. (1988). The interaction of domain-specific and strategic knowledge in academic performance. *Review of Educational Research, 58*(4), 75–404. https://doi.org/10.3102/00346543058004375.

Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2010). Does discovery-based instruction enhance learning? *Journal of Educational Psychology, 103*(1), 1–18. https://doi.org/10.1037/a0021017.

Baroody, A. J., & Dowker, A. (2003). *The development of arithmetic concepts and skills: Constructing adaptive expertise.* Mahwah, NJ: Erlbaum.

Baroody, A. J., Lai, M.-L., & Mix, K. S. (2006). The development of young children's number and operation sense and its implications for early childhood education. In B. Spodek, & O. N. Saracho (Eds.). *Handbook of research on the education of young children* (pp. 187–221). Mahwah, NJ: Erlbaum.

Bauer, D. J., & Hussong, A. (2009). Psychometric approaches for developing commensurate measures across independent studies: Traditional and new models. *Psychological Methods, 14*(2), 101–125. https://doi.org/10.1037/a0015583.

Biddlecomb, B., & Carr, M. (2011). A longitudinal study of the development of mathematics strategies and underlying counting schemes. *International Journal of Science and Mathematics Education, 9*, 1–24.

Darrell Bock, R. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika, 37*(1), 29–51. https://doi.org/10.1007/BF02291411.

Carnine, D. W., Jitendra, A. K., & Silbert, J. (1997). A descriptive analysis of mathematics curricular materials from a pedagogical perspective: A case study of fractions. *Remedial and Special Education, 18*(2), 66–81.

Carpenter, T. P., Fennema, E. H., Franke, M. L., Levi, L., & Empson, S. B. (2014). *Children's mathematics: Cognitively guided instruction* (2nd ed.). Portsmouth, NH: Heinemann.

Carpenter, T. P., Franke, M. L., Jacobs, V. R., Fennema, E. H., & Empson, S. B. (1998). A longitudinal study of invention and understanding in children's multidigit addition and subtraction. *Journal for Research in Mathematics Education, 29*(1), 3–20. https://doi.org/10.2307/749715.

Carr, M., Jessup, D. L., & Fuller, D. (1999). Gender differences in first-grade mathematics strategy use: Parent and teacher contributions. *Journal for Research in Mathematics Education, 30*, 20–46.

Choppin, J. (2011). The impact of professional noticing on teachers' adaptations of challenging tasks. *Mathematical Thinking and Learning, 13*, 175–197.

Clark, R. E., Kirschner, P. A., & Sweller, J. (2012). Putting students on the path to learning: The case for fully guided instruction. *American Educator, 36*(1), 6–11.

Clarke, B., Doabler, C. T., Nelson, N. J., & Shanley, C. (2014). Effective instructional strategies for kindergarten and first-grade students at risk in mathematics. *Intervention in School and Clinic*. https://doi.org/10.1177/1053451214560888.

Clements, D. H. (2007). Curriculum research: Toward a framework for 'research-based curricula. *Journal for Research in Mathematics Education, 38*(1), 35–70. https://doi.org/10.2307/30034927.

Clements, D. H., Agodini, R., & Harris, B. (2013). Instructional practices and student math achievement: Correlations from a study of math curricula. Retrieved from NCEE (National Center for Education Evaluation and Regional Assistance) website: http://ies.ed.gov/ncee/pubs/20134020/.

Clements, D. H., & Sarama, J. (2007/2013). *Building Blocks, Volumes 1 and 2.* Columbus, OH: McGraw-Hill Education.

Clements, D. H., & Sarama, J. (2014). *Learning and teaching early math: The learning trajectories approach* (2nd ed.). New York, NY: Routledge.

Clements, D. H., Sarama, J., & Liu, X. (2008). Development of a measure of early mathematics achievement using the Rasch model: The Research-based Early Maths Assessment. *Educational Psychology, 28*(4), 457–482. https://doi.org/10.1080/01443410701777272.

Clements, D. H., Sarama, J., Spitler, M. E., Lange, A. A., & Wolfe, C. B. (2011). Mathematics learned by young children in an intervention based on learning trajectories: A large-scale cluster randomized trial. *Journal for Research in Mathematics Education, 42*(2), 127–166. https://doi.org/10.5951/jresematheduc.42.2.0127.

Clements, D. H., Sarama, J., Wolfe, C. B., & Day-Hess, C. A. (2008/2019). *REMA—Research-based Early Mathematics Assessment.* Denver, CO: Kennedy Institute, University of Denver.

Clements, D. H., Sarama, J., Wolfe, C. B., & Spitler, M. E. (2013). Longitudinal evaluation of a scale-up model for teaching mathematics with trajectories and technologies: Persistence of effects in the third year. *American Educational Research Journal, 50*(4), 812–850. https://doi.org/10.3102/0002831212469270.

Confrey, J. (2019). A synthesis of research on learning trajectories/progressions in mathematics. Retrieved from Organisation for Economic Co-operation and Development website: http://www.oecd.org/education/2030/A-Synthesis-of-Research-on-Learning-Trajectories-Progressions-in-Mathematics.pdf.

Coyle, T. R. (2001). Factor analysis of variability measures in eight independent samples of children and adults. *Journal of Experimental Child Psychology, 78*, 330–358.

Dinsmore, D. L. (2017). Toward a dynamic, multidimensional research framework for strategic processing. *Educational Psychology Review, 29*(2), 235–268. https://doi.org/10.1007/s10648-017-9407-5.

Dumas, D. (2019). Strategic processing within and across domains of learning. In D. L. Dinsmore, L. K. Fryer, & M. M. Parkinson (Eds.). *Handbook of strategies and strategic processing: Conceptualization, intervention, measurement, and analysis*. New York, NY: Routledge.

Fennema, E. H., Carpenter, T. P., Franke, M. L., Levi, L., Jacobs, V. R., & Empson, S. B. (1996). Mathematics instruction and teachers' beliefs: A longitudinal study of using children's thinking. *Journal for Research in Mathematics Education, 27*, 403–434.

Franke, M. L., Kazemi, E., & Battey, D. (2007). Mathematics teaching and classroom practice. In F. K. Lester (Vol. Ed.), *Second handbook of research on mathematics teaching and learning: Vol. 1*, (pp. 225–256). New York, NY: Information Age Publishing.

Fuchs, L. S., Powell, S. R., Seethaler, P. M., Cirino, P. T., Fletcher, J. M., Fuchs, D., & Hamlett, C. L. (2010). The effects of strategic counting instruction, with and without deliberate practice, on number combination skill among students with mathematics difficulties. *Learning and Individual Differences, 20*(2), 89–100. https://doi.org/10.1016/j.lindif.2009.09.003.

Gersten, R. (1985). Direct instruction with special education students: A review of evaluation research. *The Journal of Special Education, 19*, 41–58.

Gersten, R., Beckmann, S., Clarke, B., Foegen, A., Marsh, L., Star, J. R., & Witzel, B. (2009). Assisting students struggling with mathematics: Response to Intervention (RtI) for elementary and middle schools (NCEE 2009-4060). Retrieved from National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education website: http://ies.ed.gov/ncee/wwc/publications/practiceguides/.

Heasty, M., McLaughlin, T. F., Williams, R. L., & Keenan, B. (2012). The effects of using direct instruction mathematics formats to teach basic math skills to a third grade student with a learning disability. *Academic Research International, 2*(3), 382–387.

Henry, V. J., & Brown, R. S. (2008). First-grade basic facts: An investigation into teaching and learning of an accelerated, high-demand memorization standard. *Journal for Research in Mathematics Education, 39*(2), 153–183.

Hickendorff, M., van Putten, C. M., Verhelst, N. D., & Heiser, W. J. (2010). Individual differences in strategy use on division problems: Mental versus written computation. *Journal of Educational Psychology, 102*(2), 438–452. https://doi.org/10.1037/a0018177.

Jacobs, V. R., & Empson, S. B. (2016). Responding to children's mathematical thinking in the moment: An emerging framework of teaching moves. *ZDM Mathematics Education, 48*, 185–197. https://doi.org/10.1007/s11858-015-0717-0.

Kam, M. (2017). Fostering students' mathematics communication in measurement: A grade 3 teacher's perceptions and practices. (Doctoral Dissertation), Ontario Institute for Studies in Education of the University of Toronto, Toronto, Canada. Retrieved from https://tspace.library.utoronto.ca/bitstream/1807/79024/1/Kam_Mimi_201706_PhD_thesis.pdf.

Kamii, C., & Dominick, A. (1998). The harmful effects of algorithms in grades 1–4. In L. J. Morrow, & M. J. Kenney (Eds.). *The teaching and learning of algorithms in school mathematics* (pp. 130–140). Reston, VA: National Council of Teachers of Mathematics.

Kerkman, D. D., & Siegler, R. S. (1993). Individual differences and adaptive flexibility in lower-income children's strategy choices. *Learning and Individual Differences, 5*(2), 113–136.

Klein, A., Starkey, P., & Wakeley, A. (2000). *Child Math Assessment: Preschool Battery (CMA).* Berkeley, CA: University of California, Berkeley.

Lemaire, P., & Siegler, R. S. (1995). Four aspects of strategic change: Contributions to children's learning of multiplication. *Journal of Experimental Psychology: General, 124*(1), 83–97.

Magliano, J. P., Trabasso, T., & Graesser, A. (1999). Strategic processing during comprehension. *Journal of Educational Psychology, 91*(4), 615–629. https://doi.org/10.1037/0022-0663.91.4.615.

McNeal, B. (1995). Learning not to think in a textbook-based mathematics class. *Journal of Mathematical Behavior, 14*(2), 205–234.

Minsky, M. (1986). *The society of mind.* New York, NY: Simon and Schuster.

Murata, A., & Fuson, K. C. (2006). Teaching as assisting individual constructive paths within an interdependent class learning zone: Japanese first graders learning to add using 10. *Journal for Research in Mathematics Education, 37*(5), 421–456. https://doi.org/10.2307/30034861.

Naglieri, J. A., & Johnson, D. (2000). Effectiveness of a cognitive strategy intervention in improving arithmetic computation based on the PASS theory. *Journal of Learning Disabilities, 33*(6), 591–597. https://doi.org/10.1177/002221940003300607.

National Research Council (2007). *Taking science to school: Learning and teaching sciences in grades K-8.* Washington, DC: National Academies Press.

Nering, M. L., & Ostini, R. (2011). *Handbook of polytomous item response theory models.* New York, NY: Taylor & Francis.

Paccagnella, O. (2006). Centering or not centering in multilevel models? The role of the group mean and the assessment of group effects. *Evaluation Review, 30*(1), 66–85. https://doi.org/10.1177/0193841X05275649.

Pang, J. S., & Kim, J. W. (2018). Characteristics of Korean students' early algebraic thinking: A generalized arithmetic perspective. In C. Kieran (Ed.). *Teaching and learning algebraic thinking with 5- to 12-year-olds: The global evolution of an emerging field of research and practice* (pp. 141–165). Cham: Springer International Publishing.

Paulos, J. A. (1991). *Beyond numeracy: Ruminations of a numbers man.* New York, NY: Alfred A. Knopf.

Peterson, P. L., Carpenter, T. P., & Fennema, E. H. (1989). Teachers' knowledge of students' knowledge in mathematics problem solving: Correlational and case analyses. *Journal of Educational Psychology, 81*, 558–569.

Plewis, I. (1989). Comment on "centering" predictors in multilevel analysis. *Multilevel Modelling Newsletter, 1(3), 6*, 11.

Powell, S. R., & Fuchs, L. S. (2015). Intensive intervention in mathematics. *Learning Disabilities Research & Practice, 30*(4), 182–192. https://doi.org/10.1111/ldrp.12087.

Pressley, M., & Harris, K. R. (2006). Cognitive strategies instruction: From basic research to classroom instruction. In P. A. Alexander, & P. H. Winne (Eds.). *Handbook of educational psychology* (pp. 265–286). Mahwah, NJ: Lawrence Erlbaum Associates.

Raudenbush, S. W. (1989). "Centering" predictors in multilevel analysis: Choices and consequences. *Multilevel Modelling Newsletter, 1*(2), 10–12.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park, CA: Sage.

Rhodes, K. T., Lukowski, S., Branum-Martin, L., Opfer, J., Geary, D. C., & Petrill, S. A. (2019). Individual differences in addition strategy choice: A psychometric evaluation. *Journal of Educational Psychology, 111*(3), 414–433. https://doi.org/10.1037/edu0000294.

Rockwell, S. B., Griffin, C. C., & Jones, H. A. (2011). Schema-based strategy instruction in mathematics and the word problem-solving performance of a student with autism. *Focus on Autism and Other Developmental Disabilities, 26*, 87–95.

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

Sarama, J., Clements, D. H., Wolfe, C. B., & Spitler, M. E. (2012). Longitudinal evaluation of a scale-up model for teaching mathematics with trajectories and technologies. *Journal of Research on Educational Effectiveness, 5*(2), 105–135 10.1080119345747.2011,627980.

Sarama, J., Lange, A., Clements, D. H., & Wolfe, C. B. (2012). The impacts of an early mathematics curriculum on emerging literacy and language. *Early Childhood Research Quarterly, 27*(3), 489–502. https://doi.org/10.1016/j.ecresq.2011.12.002.

Saxton, M., & Cakir, K. (2006). Counting-on, trading and partitioning: Effects of training and prior knowledge on performance on base-10 tasks. *Child Development, 77*(3), 767–785.

Sherin, B., & Fuson, K. C. (2005). Multiplication strategies and the appropriation of computational resources. *Journal for Research in Mathematics Education, 36*(4), 347–395.

Siegler, R. S. (1993). Adaptive and non-adaptive characteristics of low income children's strategy use. In L. A. Penner, G. M. Batsche, H. M. Knoff, & D. L. Nelson (Eds.). *Contributions of psychology to science and mathematics education* (pp. 341–366). Washington, DC: American Psychological Association.

Siegler, R. S. (1994). Cognitive variability: A key to understanding cognitive development. *Behaviometrika, 3*, 1–50.

Siegler, R. S. (1995). How does change occur: A microgenetic study of number conservation. *Cognitive Psychology, 28*, 255–273. https://doi.org/10.1006/cogp.1995.1006.

Siegler, R. S. (2006). Microgenetic analyses of learning. In W. Damon, R. M. S. E. Lerner, D. Kuhn, & R. S. V. E. Siegler (Eds.). *Handbook of child psychology: Volume 2: Cognition, perception, and language* (pp. 464–510). (6th ed.). Hoboken, NJ: John Wiley & Sons.

Steffe, L. P. (1994). Children's multiplying schemes. In G. Harel, & J. Confrey (Eds.). *The development of multiplicative reasoning in the learning of mathematics* (pp. 3–39). Albany, NY: SUNY Press.

Stein, M. K., Engle, R. A., Smith, M. S., & Hughes, E. K. (2008). Orchestrating productive mathematical discussions: Five practices for helping teachers move beyond show and tell. *Mathematical Thinking and Learning, 10*(4), 313–340.

Thornton, C. A. (1979). See Emphasis on thinking strategies in basic fact instruction-a response. *Journal for Research in Mathematics Education, 10*(3), 233–234.

Torbeyns, J., Verschaffel, L., & Ghesquière, P. (2005). Simple addition strategies in a first-grade class with multiple strategy instruction. *Cognition and Instruction, 23*, 1–21.

Verschaffel, L., Torbeyns, J., De Smedt, B., Luwel, K., & Dooren, W. V. (2007). Strategy flexibility in children with low achievement in mathematics. *Educational & Child Psychology, 24*(2), 16–27.

Wansart, W. L. (1990). Learning to solve a problem: A microanalysis of the solution strategies of children with learning disabilities. *Journal of Learning Disabilities, 23*(164–170), 184.

Weiland, C., Wolfe, C. B., Hurwitz, M. D., Clements, D. H., Sarama, J., & Yoshikawa, H. (2012). Early mathematics assessment: Validation of the short form of a prekindergarten and kindergarten mathematics measure. *Educational Psychology, 32*(3), 311–333. https://doi.org/10.1080/01443410.2011.654190.

What Works Clearinghouse. (2017). Self-regulated strategy development. Retrieved from What Works Clearinghouse website: https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc_srsd_111417.pdf.

Wright, B. D., & Stone, M. H. (1979). *Best test design: Rasch measurement.* Chicago: MESA Press.

Wu, H.-H. (2011). *Understanding numbers in elementary school mathematics.* Providence, RI: American Mathematical Society.