Journal of Mathematical Behavior 76 (2024) 101178

Contents lists available at ScienceDirect

Journal of Mathematical Behavior

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

# Does use of a hypothetical learning progression promote learning of the cardinal-count concept and give-*n* performance?

Arthur J. Baroody [a,*], Douglas H. Clements [b], Julie Sarama [c]

[a] *University of Illinois at Urbana-Champaign, Department of Curriculum & Instruction, College of Education, Champaign, IL 61820, United States*
[b] *University of Denver, Katherine A. Ruffatto Hall Rm. 152/154, 1999 East Evans Avenue, Denver, CO 80208-1700, United States*
[c] *University of Denver, Katherine A. Ruffatto Hall Rm. 160, 1999 East Evans Avenue, Denver, CO 80208-1700, United States*

ARTICLE INFO

*Keywords:*
Cardinality
Counting development
Early childhood
Give-*n* task
Hypothetical learning progressions

ABSTRACT

The general aim of the research was to conduct a rare test of the efficacy of hypothetical learning progressions (HLPs) and a basic assumption of basing instruction on HLPs, namely teaching each successive level is more efficacious than skipping lower levels and teaching the target level directly. The specific aim was evaluating whether counting-based cardinality concepts unfold in a stepwise manner. The research involved a pretest—delayed-posttest design with random assignment of 14 preschoolers to two conditions. The experimental intervention was based on an HLP for cardinality development (first promoting levels that presumably support and are necessary for the target level and then the target knowledge). The active-control treatment entailed a Teach-to-Target approach (first promoting irrelevant cardinality knowledge about recognizing written numbers and then directly teaching the same target-level goals with the same explicit instruction and similar games). A mix of quantitative and qualitative analyses indicated HLP participants performed significantly and substantially better than Teach-to-Target participants on target-level concept and skill measures. Moreover, the former tended to make sensible errors, whereas the latter generally responded cluelessly.

## 1. Introduction

A learning progression (LP) involves a theoretically- and research-based sequence of knowledge levels. A hypothetical LP (HLP) refers to a provisional and relatively fine-grained LP used to guide instruction (Lehrer & Schauble, 2015; see, e.g., the progression detailed in Table 1). A hypothetical learning trajectory (HLT) is a special kind of HLP, because in addition to a fine-grained LP, it includes instructional goals and theoretically and empirically based instructional tasks for achieving each level of an HLP (Lobato & Walters, 2017; e.g., see the instructional details of the experimental condition in Appendix C of the Supplementary materials). As our discussions of the theoretical basis (e.g., assumptions), existing research, and applications to curriculum and instruction apply to both HLPs and HLTs, we use the general term HLP.

Although reformers have hailed HLPs as an important tool in improving mathematics education, direct research on their assumptions and efficacy is limited (Frye et al., 2013; Lobato & Walters; Shavelson & Karplus, 2012). The general aim of the present research was to contribute to the small and contradictory evidence supporting HLPs by rigorously evaluating a basic assumption for their use. This general aim was evaluated in the context of the specific aim of appraising whether key counting-based cardinality

* Corresponding author.
*E-mail addresses:* baroody@illinois.edu (A.J. Baroody), Douglas.Clements@du.edu (D.H. Clements), Julie.Sarama@du.edu (J. Sarama).

https://doi.org/10.1016/j.jmathb.2024.101178
Received 23 October 2023; Received in revised form 4 July 2024; Accepted 10 August 2024

concepts unfold in a stepwise manner.

## 2. General issue: is a basic assumption of HLP-based instruction valid?

### 2.1. Purported value of HLPs

Proponents of the traditional, didactic "Teach-to-Target" approach argue that immediately teaching the target level by explicitly providing definitions and procedures and drilling targeted knowledge is more efficient and mathematically rigorous than slowly moving through a series of levels, especially if it involves discovery learning (see Bereiter, 1986; Clark et al., 2012; Kirschner et al., 2006; Rosenshine, 2009, 2012; Wu, 2011). For instance, students presumably learn accurate target-level *and* lower-level knowledge in less time. An example of this approach is the "worked examples" method—explicitly describing and illustrating how to solve a new type of problem, including the conceptual rationale for each step (Clark et al., 2006; Renkl, 2005; Schwonke et al., 2009). Such instruction too often is implemented with little consideration for building on prior knowledge and results in little or no understanding of conceptual prerequisites for target instruction or the target knowledge itself (Baroody & Pellegrino, 2023; Clements et al., 2021, 2023).

Those interested in educational reform have long recommended gradually building on prior knowledge as a means for overcoming the limitations of rote memorization engendered by traditional, didactic instruction. For example, in "Talk to Teachers," the eminent psychologist William James (1958) advocated promoting meaningful memorization: "Effort should not be so much to *impress* and *retain* [new information] as to *connect* it with something already there" (pp. 101–102). Similarly, Piaget (1964, p. 18) argued that "the fundamental relation from the point of view of pedagogical … application" is not associations, but assimilation: "the integration of any sort of reality into [an existing] structure." Since the late 20th century, reformers have become increasingly interested in supporting such connections by developing, promoting, and using HLPs—building on a sequence of prior knowledge levels (Resnick & Ford, 1981; Shavelson & Karplus, 2012; Vygotsky, 1934/1986).

HLPs are deemed valuable because they can highlight developmentally appropriate and important goals and help focus instructional efforts on them. They also underscore how children typically develop, the need to consider what they must already know to make progress, and what level of instruction is within their comprehension (cf. Vygotsky's Zone of Proximal Development). HLPs,

**Table 1**
A possible HLP of key aspects of pre-counting and counting-based cardinal number knowledge and possible operational definitions.

| Possible HLP Level | Possible Task |
|---|---|
| *Levels 1A & 1B. Pre-counting (subitizing-based) number recognition & set creation:* *Subitizing* entails recognizing, without counting, the total number of items in a collection and labeling it with an appropriate number word (Kaufman et al., 1949). Such small-number recognition (Level 1A) serves as a scaffold for constructing subsequent levels (Benoit et al., 2004; Schaeffer et al., 1974; Von Glasersfeld, 1982). For instance, without counting, it may enable children to "see" when they have put out a requested number of objects from a larger pile of objects (Level 1B). | *Number recognition (Level 1A)*: reliably indicate the total of small sets on a how-many task without counting. *Set creation (Level 1B):* reliably produce a small set on a give-*n* task (e.g., "give-me three blocks") without counting. |
| *Transition to meaningful counting:* Children often learn to count in a one-to-one fashion by rote—without realizing its purpose is to determine the total number of items in a collection. Fuson (1988) observed that, frequently, they subsequently learned to recognize that repeating the last number word used in the counting process is an acceptable response to a how-many question. Put differently, they learn a non-meaningful "last-word rule," which creates the appearance of knowing how many. | *Last-word rule*: reliably responds with the last number word used in counting a set on a how-many task (without understanding it represents the total) |
| *Level 2. Meaningful counting:* In time, they construct the basis of meaningful object counting namely, the cardinality principle (CP): the understanding that the last number-word used in the counting process has special significance because it represents the total number of items in a collection (Level 2). Fuson (1988) called the CP the "count-cardinal transition," because a **count** serves to determine the **cardinal** value of a collection. Modeling the CP with small subitizable collections can facilitate discovering the CP—literally help children see that the last number word in a count matches a collection's total number (Paliwal & Baroody, 2020). | *CP or count-cardinal transition*: reliably responds with the last number word used in counting a set on a how-many task (with the understanding it represents the total—as independently confirmed by a task the last-word rule cannot be used) |
| *Level 3. CP Applications (Cardinal-Number Constancy Concepts):* Subitizing experiences and an understanding of the CP underlies constancy concepts (Fuson, 1988; Schaeffer et al., 1974) such as counting-based conservation of cardinal identity: recognizing that a counting-generated cardinal label is still applicable even if the appearance of a collection is changed. | *CP application:* reliably count a larger set (e.g., 6 blocks), indicate the total, and after the blocks are re-arranged indicate the same total. |
| *Level 4. Counting-out fluency and understanding:* Fuson (1988) hypothesized the CP serves as a developmental prerequisite for the concept she called the "cardinal-count transition": understanding that the **cardinal** label of a collection determines what the last number word would be *if* the collection was **counted**. The latter concept, which entails a (counting-based) word-to-quantity mapping, is the inverse of the former, which involves a (counting-based) quantity-to-word mapping. | *Counting-out fluency*: reliably counting out a requested number of items (beyond the subitizing range). *Counting-out understanding (cardinal-count transition)*: reliably indicate that, for example, "give six items," the counting-out process should stop when the count reaches "six." |

**Table 2**
A description of the testing tasks.

| Number: Name (Cardinality Level) | Materials | Procedure | Scoring |
|---|---|---|---|
| Task #1: Verbal counting | iPad | A child was encouraged to verbally count as an avatar pointed to each of 10 blocks. | The child's high count was the last correctly stated number before an error and determined the number of blocks that remained in the child's tower after the tower was shaken. It was compared to an avatar's tower, which always had fewer blocks. |
| Task #2: Subitizing-based how many (Level 1A) | iPad | The task involved 1–4 fish swimming across a screen for 2 s. A child was asked: "How many fish did you see?" For each number 1–4, there were up to three trials. If the child was incorrect or correct on the first two trials of a number, testing stopped for that number. If the child was correct on only one of the first two trials, a make-up or tie-breaking trial was administered. | For each number 1–4, the criterion of success was two correct responses and no more than one overuse of the number word (e.g., for $n = 2$, labeling a collection of three as "two"). The minimum criterion for inclusion in the study was a 1-knower (successfully identifying 1) and at least transitioning to the 2-knower level (i.e., used "two" to label collections of two or more—to indicate "more than one"). |
| Task #3A & #3B: Give 3 and 4, respectively—subitizing-based set creation (Level 1B) | Six bean bags of one color and six of another; cardboard box (which served as a target) | To demonstrate the procedure, a child was asked to give the trainer 2 bean bags from a pile of six bean bags so that the tester could take her turn. The child was asked to take *n* bean bags from a pile of differently colored bean bags (Test Trial 1). A second game created an opportunity Test Trials 2 and 3. | For each number, a child's diagnostic score could range from 0 to 3 correct responses. In (the rare) cases where a child put out or counted all the bean bags initially but was correct after prompted, the trial was scored as correct. |
| Task #4A: Counting-based how-many with linear arrays: CP-based one-to-one counting that minimizes the procedural demand of keeping track (Level 2) or last-word-rule-based) one-to-one counting (transition to Level 2) | iPad displays of linear arrays of stars ½-in. apart. The practice trial involved a display of 2 stars; Trial 1, 5 stars; Trial 2, 7 stars; and Trial 3, 6 stars. | After the practice trial was administered, the three test trials were then administered in the order. For Step 1 (assessment of one-to-one counting), a child was asked to count a row of stars. For Step 2 (assessment of CP or last-word rule): After an avatar hid the stars, the child was asked how many stars are hidden. | Scoring for Step 1 (one-to-one counting): For each trial, the child was scored for diagnostic purposes as (a) correct, (b) made one or two minor counting errors but honored the one-to-one principle, or (c) incorrect. A minor counting error could involve a single (a) sequence error (e.g., counting a collection of six with "one, two, three, four, five, eight"), (b) one-to-one correspondence error (e.g., pointing to the fourth item of a collection of six but failing to tag it with a number word and ending the count with "five"), or (c) keeping-track error (e.g., skipping over the fourth item and counting a collection of six as "five" or losing track of which items had already been count and recounting an item a second time and counting a collection of five items as "six"). For screening purposes, children were considered to understand the one-to-one counting principle if they were correct or honored the principle on most trials for Tasks 4A and 4B. Scoring for Step 2 (CP): Correct (scored as 1) was defined as a response to the how-many question that matched the last number used in a count that honored the one-to-one counting principle (i.e., was either accurate or involved only one or two minor slips). Incorrect (scored as 0) was defined as responding otherwise, including |

(*continued on next page*)

**Table 2** (*continued*)

| Number: Name (Cardinality Level) | Materials | Procedure | Scoring |
|---|---|---|---|
| | | | simply repeating the original count or attempting to count a mental image of the items. Scores for Task 4A or 4B could range from 0 to 3 correct; scores for Task 4A and 4B combined, 0 to 6. For screening purposes, a child with a combined CP score of more than three correct was excluded from study. |
| Task #4B:Counting-based how-many task with haphazard arrays:CP-based one-to-one counting that minimizes the procedural demand of keeping track (Level 2: meaningful 1-to-1 counting) or last-word-rule-based) one-to-one counting (transition to Level 2) | Red (poker) chips, 1.5-in. in diameter, were secured to a mat using Velcro to ensure each child saw the same haphazard arrangement about 1.5 in. apart (on average). The Practice Trial involved 3 chips. Test Trials 1, 2, and 3 involved 6, 5, and 7 chips, respectively. | Identical to Task 4A, except that the word 'chips' was substituted for 'stars.' | Identical to Task 4A. |
| Task #5: Application of the CP without the "key words" "how many?" to avoid invoking the last-word rule (Level 2) | Seven 1- × 1-in. white plastic blocks and a screen. | After a practice trial, three test trials ($n = 6, 7, 5$) were administered using the instructions: "I'm hiding some blocks behind the screen. The clue is [tester verbally counts to *n*]. Can you guess the number of blocks I've hidden?" No feedback was given on test trials. | A correct response entailed stating the cardinal value of the collection. Repeating the count, stating a number other than the cardinal value, responding "I don't know," or making an irrelevant response was scored as incorrect. One point was awarded for each correct response; scores could range from 0 to 3. |
| Task #6A & #6B: Give-*n* with 5 to 7 items—knowledge about the counting-out procedure and the cardinal-count transition, which provides an understanding of why the specified number indicates when to stop (Level 4: meaningful counting-out) | iPad for virtual displays of items to be counted out (6A) or cookie monster muppet and chips (6B). | To introduce the practice trial, a child was asked to give Cookie Monster three items. The test trials involved request to give five donuts, six cookies, and seven pieces of pizza for Task 6A and take five, seven, and six cookies (chips) for Task 6B. If a child responded incorrectly on a trial, the child was given a second chance. | Credit (1 point) for was awarded on a trial if the child (a) correctly counted out the requested amount or (b) made a spontaneous correction (e.g., asked to give six cookies, initially violated the concept by counting out ten cookies, said "oops" [recognized the violation], counted six of the cookies, and pushed the rest back [correctly applied the concept]). Partial credit (0.5 points) was granted if a child made a minor counting error but otherwise followed the procedure that honored the cardinal-count concept:<br>• was within one of the requested number if the child used the standard count sequence);<br>• ended the verbal count with the requested number); and<br>• stopped the counting-out process at the requested number.<br>Partial credit (0.5 points) was also awarded for correctly counting out the requested number on a make-up trial. Scores on Task 6A (or 6B) could range from 0 to 3; scores for Task 6A and 6B combined, 0 to 6. There was 91.1 % inter-rater agreement for assigning a numerical score (0, 0.5, or 1) to Task 6A and 6B trials; there was an 88.8 % agreement for assigning a particular strategy used on a trial. All disagreements were easily resolved by a follow-up discussion. For screening purposes, a child with a combined score of more than three was excluded from study, because |

(*continued on next page*)

**Table 2** (*continued*)

| Number: Name (Cardinality Level) | Materials | Procedure | Scoring |
|---|---|---|---|
| | | | such children had achieved the target level or were well on their way to doing so. |
| Task #7: Stop at *n* with 5–7 items—a direct measure of the cardinal-count transition (conceptual basis of Level 4) | Cookie Monster muppet; 10 blocks | A trial began with the tester telling Cookie Monster to give the tester n blocks. The first and second trial involved asking for 6 and 7 blocks, respectively. For each trial, the Muppet counted all 10 blocks UNLESS stopped by the child. The first make-up trial involved a request for 5; the second, a request for 8. | 1 point was awarded for stopping the counting-out process at requested number on an initial test trial; 0.5 point was awarded for doing so on a make-up trial; and 0 points were scored otherwise. Scores could range from 0 to 2. There was a 100 % inter-rater agreement on the straightforward scoring of the task. |

then, spotlight the need for formative assessment—testing done before or during instruction to gauge a student's existing knowledge and thinking so as to tailor instruction to promote progress (assessment *for* learning)—as opposed to summative assessment—testing done after instruction to mastery of the instruction's aims (assessment *of* learning). Put differently, formative assessment serves to determine a child's existing developmental level, so that instruction can build on this knowledge to meaningfully and effectively teach the next level. For these reasons and more, researchers, educators, and policy makers have recommended HLPs as a useful tool for teachers in helping them to understand, promote, and assess children's mathematical learning (Baroody, 2016a, 2016b; Baroody & Pellegrino, 2023; Frye et al., 2013; Maloney et al., 2014; Sarama & Clements, 2009; Simon, 1995).

HLPs may be particularly important for early childhood mathematics education for several related reasons. One is that early childhood educators too often have minimal, if any, training on mathematics development and education (Institute of Medicine (IOM) and National Research Council (NRC), 2015). As a result, they frequently underestimate young children's (informal) mathematical knowledge, mechanically teach the lessons specified in a curriculum guide or textbook, and focus on basic numeracy content, which many, or even most, children have already learned (Engel et al., 2013; Kilday et al., 2012). Indeed, because of a negative disposition towards mathematics instructions, many early childhood teachers do not set any mathematical goals, use any mathematical curriculum or resources, and rely on (hit-or-miss) opportunities that emerge from children's play or routine activities (Balfanz, 1999; Li et al., 2015). The pedagogical knowledge and learning expectations of teachers of academically at-risk children are particularly unlikely to foster numeracy (Ferguson, 1998; Lee & Ginsburg, 2007).

### 2.2. Existing evidence

Although HLPs are often recommended as a valuable educational tool, little direct and rigorous empirical evidence has evaluated their underlying assumptions and efficacy. Most research in this area has focused on empirically validating the developmental levels of LPs by using a cross-sectional methodology or tracking the progress of individuals over time. Although research has shown interventions that have HLPs as a component are efficacious in promoting numeracy in realistic school settings, this research confounded the use of HLPs with other factors, such as specific instructional activities or professional development (Frye et al., 2013). Thus, little or no prior research directly or systematically examined the unique contribution or assumptions of HLP-based instruction (Frye et al., 2013; Lobato & Walters, 2017).

To address these gaps in the research, the authors designed a series of 10 experiments to examine the unique contributions of HLPs to mathematics learning (see Baroody et al., 2023, for a summary of these experiments). Two assumptions at the core of the HLP instructional approach were rigorously tested.

1. Progressively teaching one level above a child's existing level on an HLP is more efficacious than skipping a level and directly teaching to the level.
2. Presenting instructional activities in the developmental order hypothesized by an HLP is more efficacious than instruction using the same not so ordered activities.

The present research focused on evaluating the first assumption.

To test Assumption 1, six experiments served to compare HLP-based and Teach-to-Target (skip-level) instruction. Half of these comparisons significantly favored the former: Clements et al. (2019) involved shape composition and Clements et al. (2020) and Clements et al. (2021) focused on early arithmetic. A possible limitation of Clements et al. (2021) experiment, though, was a testing effect due to checking experimental, but not skip-level, participants' progress during the training. The other three (unpublished) experiments—the first two involving cardinality and other aspects of numeracy and the third focused exclusively on cardinality (effectively the pilot study for the present research)—showed no statistically significant difference between the interventions.

Initial efforts to test Assumption 1 in the domain of cardinality were not successful in no small part because they involved a novel variation of a true experiment, which took trial and error to hone. To test whether the effect of the target cardinality instruction

**Table 3**
Summary of the timing of the testing and training and the tasks used for the pretest and posttest.

| | | | |
|---|---|---|---|
| PRETESTING (4 sessions over 3 weeks) | | | |
| Session 1 | Session 2 | Session 3 | Session 4 |
| iPad familiarization (2 games) | Task 3A (give 2) | Task 3B (give 3) | Task 3C (give 4) |
| Task 1 (verbal counting) | Task 4A (how many) | Task 4B (how many) | Task 5 (CP app) |
| Task 2 (subitizing) | Task 6A (give 5–7) | Tasks 6B (give 5–7) | Task 7 (stop at *n*) |
| HLP AND TEACH-TO-TARGET INTERVENTIONS (12 sessions [about 20 min each] over 6 weeks) | | | |
| (DELAYED) POSTTESTING (2 weeks after the interventions ended; 2 sessions over 2 weeks) | | | |
| Session 1 | | Session 2 | |
| Warm-up: Task 3B (give 3) | | Warm-up: Task 3C (give 4) | |
| Task 6A (give 5–7) | | Tasks 6B (give 5–7) | |
| - | | Task 7 (stop at *n*) | |

depended on whether or how students were prepared beforehand, participants in the HLP-based received instruction hypothesized to support the target instruction; the latter received cardinality instruction not related to the target instruction; *and* both conditions received the same final step of cardinality instruction. With this design, differences in outcomes between the conditions could be attributed to the differences in previous parts of the interventions. The issues encountered with this design and other methodological limitations and the methodological refinements of the present research are discussed in the last paragraph of Section 3.

## 3. Specific issue: should cardinality development be fostered in steps or directly?

Discussed next is a more detailed theoretical rationale for our three previous (unsuccessful) cardinality studies and the present cardinality study and additional evidence bearing on different hypotheses of cardinality development.

### 3.1. Different perspectives on cardinality development

Some researchers hypothesize that the knowledge to successfully count out a requested number of items must build on earlier levels of cardinality knowledge and requires HLP-based instruction (Baroody & Lai, 2022; Frye et al., 2013; Fuson, 1988). Others presume a single level of counting-based cardinality knowledge (Brousseau, 1997; Sarnecka & Carey, 2008).

#### 3.1.1. Stepwise cardinality development

A possible HLP for the stepwise development of cardinality knowledge is described in Table 1. Briefly, the pre-counting competence of subitizing-based number recognition—quickly identifying the total of a small set and labeling it with a number word (Level 1A)—provides a basis for creating small sets of a specified number (Level 1B) and facilitates the achievement of higher levels. Meaningful object counting (Level 2) requires learning the cardinality principle (CP) or what Fuson (1988) calls the “count-cardinal transition”: understanding that the last number-word used in the counting process has special significance because it represents the total number of items in a collection. Levels 1 and 2 provide a basis for constructing Level 3 (cardinal constancy concepts) and Level 4 (counting-out and its underlying conceptual basis). Fuson hypothesizes that the conceptual basis for counting out is the *inverse* of the CP, namely the

**Table 4**
Pretest and posttest results by task and condition.

| Time point | Pretest | | | | | | Delayed posttest | |
|---|---|---|---|---|---|---|---|---|
| Task<br>Minimum – maximum score<br>Level | 1-1 counting (Task 4) 0–6<br>Level 2 | CP: How many (Task 4)<br>0–6<br>Level 2 | CP application (Task 5)<br>0–3<br>Level 2 | Set production (Give *n*)<br>$n = 3$ & 4 (Task 3)<br>0–6<br>Level 1B | <br>$n = 5–7$ (Task 6)<br>0–6<br>Level 4 | Stop at *n* (Task 7)<br>0–2<br>Level 4 | Set production (Give *n*)<br>$n = 5–7$ (Task 6)<br>0–6<br>Level 4 | Stop at *n* (Task 7)<br>0–2<br>Level 4 |
| **HLP Mean** | 4.286 | 0.143 | 0.357 | 1.286 | 0.357 | 0.071 | 3.714 | 1.143 |
| **(S.D.)** | (1.629) | (0.378) | (0.748) | (1.380) | (0.627) | (0.189) | (2.068) | (0.748) |
| **Teach-to-Target** | 4.642 | 0.571 | 0.500 | 0.429 | 0.214 | 0 | 0.143 | 0.125 |
| **Mean (S.D.)** | (1.215) | (0.787) | (0.866) | (0.787) | (0.567) | (0) | (0.350) | (0.354) |
| ***U*-value** | 22.5 | 17 | 24 | 15.5 | 17.5 | 21 | 4 | 3.5 |
| ***z*-score** | − 0.192 | − 0.894 | 0.000 | 1.086 | 0.831 | 0.383 | 2.556 | 2.619 |
| ***p* (2-tailed)** | .849 | .373 | 1.000 | .276 | .407 | .704 | .010* | .009* |
| **Cohen’s *d*** | − 0.248 | − 0.691 | − 0.177 | 0.763 | 0.239 | 0.531 | 2.196 | 1.740 |

* $p < .05$.

**Table 5**
Proportion of posttest responses on give-*n* task from most accurate to least accurate by condition.

| Condition | Attempts | Correct | Minor error | End-with-*n* error | Multiple counting-out errors | Counted out all | Put out (without counting) | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | some items | all available items |
| HLP | 52 | .4808 | .0577 | .1538 | .1346 | .0962 | .0192 | .0577 |
| Teach-to-Target | 70 | .0000 | .0286 | .1000 | .2286 | .0571 | .3000 | .2857 |

[1]Includes two test trials per child and make-up trials if the child was incorrect on a test trial.

"cardinal-count transition"): understanding that the cardinal label of a collection indicates what the last number word would be if it were counted. When asked to count out a specified number of items, many children initially do not stop the counting-out process at the requested number but simply count all the items available (Resnick& Ford, 1981). Fuson hypothesizes that the cardinal-count concept provides the rationale for monitoring the counting-out process and stopping it at the requested number. In brief, the key implication of the HLP in Table 1 for the present study is that the CP—the conceptual basis for Level 2 (meaningful one-to-one counting)—is necessary, but not sufficient, for learning the cardinal-count transition—the conceptual basis for Level 4 (meaningfully learning the counting-out procedure).

### 3.1.2. *Single-step cardinality development*

Brousseau's (1997) recommends an integrated approach to whole-number instruction—fostering meaningful one-to-one counting and counting-out simultaneously. On his view, a number problem should *bridge* counting a collection (Level 2 in Table 1) and counting-out an equivalent collection with different items (Level 4) by establishing a one-to-one correspondence between the two collections to verify the validity of the counting-out effort (Dorier, 2015).

In a similar vein, Sarnecka and Carey (2008) propose that meaningful one-to-one counting and counting out have the same conceptual basis, namely the CP, and all emerge together in a single developmental level. Citing Le Corre et al.'s (2006) evidence, they conclude that the earlier success on a how-many (one-to-one counting) task than on the give-*n* task (counting out a specified number of items beyond the subitizing range) is *not* due (in part) to the count-cardinal transition developing before the cardinal-count transition, as hypothesized by Fuson (1988). Instead, Sarnecka and Carey attribute the performance gap between the how-many and give-*n* tasks wholly to what Fuson (1988) called the "last-word rule" learned by rote: Respond with last number used in the counting process when asked, "How many?" This rule is effectively a transitional phase to the CP in that children recognize the last word used in the one-to-counting process has special significance because it seems to satisfy adults who ask the "how many?" However, children still do not understand the purpose of counting or that the last number word represents to the total number (cardinal value) of a counted collection. In brief, on Sarnecka and Carey's single-step, single-concept view, both meaningful one-to-one counting and counting out require children to induce only one concept, namely cardinality principle, the CP—its inverse (the cardinal-count transition), is an unnecessary construct (Barbara W. Sarnecka, personal communication, April 1, 2024).

## 3.2. *Existing evidence and its limitations*

The evidence supporting the HLP in Table 1 generally and Fuson's (1988) hypothesis that the CP (Level 2) develops before and is a necessary condition for the cardinal-count concept (Level 4) specifically is sparse and inconsistent (Baroody & Lai, 2022).

### 3.2.1. *Support for stepwise development*

Fuson (1988; Study 8.2.2) used a how-many task in combination with a follow-up task to distinguish between children who knew the last-word rule and the CP and a prediction task to directly gauge understanding of the cardinal-count concept. The latter involved, for instance, presenting a collection of six butterfly stickers and asking, "If you count the butterflies, what will you say for the last butterfly?" Fuson found that only the four preschoolers who appeared to know or discover the CP consistently demonstrated an understanding of the cardinal-count concept, whereas the remaining 22, who were last-word responders or had been taught the last-word rule, failed to do so. Though Fuson's evidence is consistent with her hypothesis, it is not conclusive. The results do not preclude the alternative conclusion that the CP and cardinal-count concepts emerge *simultaneously* (i.e., co-evolve after children construct the

**Table 6**
Proportion of posttest responses on the stop-at-*n* task from most accurate to least accurate by condition.

| Condition | Total trials attempted[a] | Most accurate <==================> Least Accurate | | | |
|---|---|---|---|---|---|
| | | Stopped count at *n* (requested number) | Off by 1 | Off by $\geq$ 2 | No effort to stop count |
| HLP | 22 | .45 | .18 | .27 | .09 |
| Teach-to-Target | 28 | .00 | .11 | .68 | .21 |

[a] Includes two test trials per child and make-up trials if the child was incorrect on a test trial.

last-word rule). Moreover, as Fuson observed, the wording of the prediction (cardinal-count) task may have been confusing to some children.

Using latent variable modeling of 3- and 4-year-olds' performance on the how-many (set-to-word) and give-*n* (word-to-set) tasks with sets of one to eight items, Mou et al. (2021) found that the best-fitting model was a bi-factor model indicating that the two tasks, though related, reflect distinct conceptual knowledge. O'Rear et al. (2024) replicated these results with somewhat older children. Although both sets of results are consistent with Fuson's (1988) hypothesis that the cardinal-count concept requires understanding of the CP, neither entailed distinguishing among accurate responses on the how-many task due to subitizing, a last-word rule, or the CP. Additionally, in the case of Mou et al. (2021), success on the how-many task was defined as accurately counting a collection, but accurate one-to-one counting emerges before CP understanding (Schaeffer et al., 1974).

Baroody and Lai (2022) found evidence that knowledge of the CP (as assessed by a combination of results on the how-many task and an CP-application task) emerged before success on the give-*n* task. This research, however, did not involve a direct measure of the cardinal-count transition and performance factors required by the give-*n* task may have underestimated the concept.

#### *3.2.2. Support for single-step development*

To date, Brousseau's (1997) recommendation of integrated whole-number instruction does not appear to have been formally and rigorously evaluated using experiments. Le Corre et al.'s (2006) evidence cited by Sarnecka and Carey (2008) to support their single-step, single-concept hypothesis does not clearly do so. Le Corre et al. presented a card with one to seven stickers and asked, "What's on this card?" (the WOC task) instead of a how-many task to gauge the CP, because the latter may be unclear, confusing, or misleading to young children and *underestimate* CP knowledge (e.g., Gelman, 1993). As the results of the WOC and give-*n* tasks did not differ significantly, LeCorre et al. concluded the tasks provide "overwhelmingly consistent pictures of what children understand about how counting represents number" (p. 151). In fact, four of the 19 children who were identified as CP-knowers on the WOC task were identified as non-CP-knowers on the give-*n* task, and the nonsignificant results may have simply been due to a lack of power (see Baroody & Lai, 2022, and Baroody et al., 2023, for a detailed analysis). Moreover, these results indicate that in more than a fifth of the cases, understanding the CP was not sufficient for success on give-*n* task with numbers beyond the subitizing range and the difference in performance cannot be attributed to simply knowing the last-word rule.

Consistent with Sarnecka and Carey's (2008) single-step, single concept view, Paliwal and Baroody (2020) found that learning the CP can promote learning of the counting-out procedure without instruction on the cardinal-count concept or the procedure itself. However, the results can also be explained by the stepwise-development hypothesis as transfer: Learning the CP promoted learning of its inverse (the cardinal-count transition) and the counting-out procedure. Moreover, as the results of Baroody and Lai (2022) and Le Corre et al. (2006) indicate, not all children who *understand* the CP are successful on the give-*n* task.

Although the authors' three unsuccessful efforts to support Assumption 1 likewise appear to support the undifferentiated view of cardinality development and contradict Fuson's (1988) hypothesis of stepwise cardinality development, these experiments had methodological limitations. A post-mortem of first unsuccessful cardinality experiment revealed that non-significant difference between the HLP-based and the Teach-to-Target interventions may have partly been due to insufficiently distinct instruction. Specifically, target-level (Level-4) activities from the Building Blocks curriculum used for both interventions involved lower-level (Level-2) competencies as well. A lack of fidelity to both the HLP and Teach-to-Target instructional protocols may have plagued both the first and second unsuccessful cardinality experiments. For example, the trainers did what naturally educators do, which was they help a child with both Level-2 and -4 competencies regardless of a child's assignment (by, e.g., unintentional facial expression or gestures). An imprecise measure of the cardinal-count concept may have obscured the results of all three previous cardinality experiments. Specifically, a give-*n* task was used as the dependent measure and this task confounds conceptual understanding of the cardinal-count transition with procedural fluency. Finally, participants selection appeared to be a factor in all three experiments. For example, in the third unsuccessful cardinality experiment, the length of the training was insufficient to help a child in the experimental condition who could not achieve the Level 1 competencies of subitizing one or two, and several children in the Teach-to-Target condition already had achieved some success with counting-out.

## 4. Aims of the present study and its design rationale

The present research served to directly test Assumption 1 by testing the hypothesis: Instruction is more efficacious if it builds successively on children's cardinal knowledge than if counting out a specified number of items and its conceptual rationale (Level 4 in Table 1) is taught directly. The domain of counting and cardinality knowledge was chosen as the context for testing Assumption 1, and counting out a specified number of items *with* understanding was chosen as the target level for three reasons:

1. Counting and cardinality knowledge is a foundation for learning a variety of informal and formal mathematical knowledge (Council of Chief State School Officers, 2010). Furthermore, the target knowledge has long been recognized as a basic competence important for everyday life and success with school mathematics (Resnick& Ford, 1981). For example, Geary et al. (2018) found that the age children achieved Level 4 is more strongly related to mathematical development and school readiness than reading achievement, controlling for intelligence and executive function.
2. An evaluation of Assumption 1 (or 2) with an HLP in which an earlier level is a *necessary* condition for a later one is more likely to provide clearcut and significant support than that with an HLP in which an earlier level is merely a *facilitative* one (Baroody et al., 2023). In the later case, even if more children in the HLP condition achieve the target knowledge, some in the control condition may also achieve it because the experimental HLP training, though helpful, is not indispensable. This may prevent achieving statistical

significance or even practical significance (as measured by effect size) and may help to account for our first two unsuccessful efforts to corroborate Assumption 1 (and several unsuccessful efforts to corroborate Assumption 2).
3. The three successful efforts to test Assumption 1 involved domains other than cardinality (Clements et al., 2020; Clements et al., 2019; Clements et al., 2021). Examining the domain of cardinality might serve to increase the generalizability of previous results.

The design of the present research avoided the methodological limitation of a prior HLP efficacy study (Clements et al., 2021) by ensuring that both the HLP and Teach-to-Target groups had the same exposure to testing on the dependent measures. It also avoided methodological pitfalls of prior cardinality studies, including the authors' three previous unsuccessful efforts to corroborate Fuson's (1988) hypothesis. Particularly important refinements over one or more previous efforts were—

1. Including a means of gauging whether correct responses on the how-many task were meaningful (i.e., the product of the CP) or not (i.e., due to the last-word rule).
2. Using a direct measure of cardinal-count transition (as well as the give-*n* task, which assesses both conceptual *and* procedural knowledge).
3. Excluding children who were not developmentally ready for the experimental intervention (i.e., who could not reliably recognize and label sets of 1).
4. Excluding children who had already achieved Level 2 or 4 to better test the hypothesis of whether skipping a level makes a difference.
5. Carefully constructing the Teach-to-Target training—including the target (Level-4) training–to avoid Level-2 (CP) instruction.

## 5. Method

### 5.1. Participants

Children were recruited from 10 morning and afternoon classes of a public preschool program, serving a predominantly working- and middle-class community in a mid-sized, midwestern U.S. city. Pretesting identified 14 eligible participants—children with (some) Level 1 knowledge but who had not achieved (a) Level 2 (i.e., one-to-one counting with CP understanding) or (b) the Level-4 target instruction (the cardinal-count transition and accurate counting-out of larger numbers). (See Appendix A in the Electronic Supplemental Materials for details regarding the selection criteria, their rationale, and resulting attrition.) Four of the seven participants in HLP group who completed the study were girls ($M = 4.44$ years, $SD = 0.66$ years, range = 3.51–5.03 years); four of the seven participants in the Teach-to-Target group were girls ($M = 3.83$ years, $SD = 0.55$ years, range= 3.23–4.76 years). The ethnic background of the HLP group was 1 Caucasian, 5 African American, and 1 Hispanic; the Teach-to-Target group, 2 Caucasian and 5 African American).

### 5.2. Tasks

The pretest consisted of game-based tasks designed to assess seven constructs. Pretesting served to identify a participant's level of verbal-based cardinality development for screening and diagnostic purposes. The posttest employed two of these tasks—counting-out 5 to 7 (Task 6) and stop at *n* (Task 7) as the dependent measure for procedural fluency and conceptual understanding, respectively. Procedural fluency—the meaningful and effective use of a routine—requires the integration of procedural and conceptual knowledge (Kilpatrick et al., 2001). In the case of counting out a specified number of items, procedural knowledge includes knowing how to implement one-to-one counting and a keeping-track strategy (a means of identifying which items have already been counted out and which have not), and conceptual knowledge entails understanding the cardinal-count transition (comprehension of why the number specified in a give-*n* task indicates when to stop the counting-out process). The stop-at-*n* task directly measures the cardinal-count transition without the procedural demands of the give-*n* task. See Table 2 for a description of the tasks and Appendix B in the Supplemental materials for additional task details, including probes/prompts.

### 5.3. Interventions

Both the HLP and Teach-to-Target interventions initially focused on the identical preliminary training—review or remedial training of subitizing-based number recognition and putting out up to four requested items (Levels 1A and 1B, respectively). In subsequent sessions, remedial subitizing efforts continued as needed for HLP participants only. This was done to mimic the fundamental assumption of HLP-based instruction that meaningful instruction should build on solid prior knowledge and the common Teach-to-Target practice of not doing so. Specially, Paliwal and Baroody (2020) found that the ability to subitize up to four can facilitate the impact of modeling the CP, which was an aspect of the HLP training but not the Teach-to-Target training. Moreover, we wanted to partially mimic HLP-based instruction that involves mastering an earlier level before starting on a higher level. Fully implementing this characteristic was not practical because of time constraints and the need to avoid providing the HLP group more instructional time than the Teach-to-Target group (i.e., creating a dosage confound). Even when earlier knowledge is taught in the Teach-to-Target approach, instruction is typically done in a lock-step manner and instruction moves on to subsequent instruction before all children have mastered more basic knowledge.

The HLP intervention next focused on the proposed developmental prerequisite, the CP (Level 2; Sessions 3–6), and—to consolidate Level-2 learning—applications of the concept (Level 3; Sessions 7 & 8). For example, the *Hidden Chips* game involved having a child count a collection of chips, the trainer covering the chips and asking the how-many question, and the child applying the CP. Feedback was provided by uncovering the collection and either announcing "correct" or modeling the concept. Meanwhile, the Teach-to-Target intervention focused on a different, but educationally valuable, aspect of cardinality: recognizing the written numbers 0 to 9 (i.e., connecting written numbers to verbal number knowledge) and relating the symbols 1 to 5 to quantities. Whereas HLP children counted dots to determine a cardinal number, Teach-to-Target participants did so by reading a written numeral.

Both interventions had the same Level-4 goals and the same dosage of the target-level instruction (Sessions 9–12): the cardinal-count transition and how to use it to count-out a collection. The two interventions involved identical and explicit conceptual training on the cardinal-count transition. This was in the form of an error-detection activity (*Good or Bad Counter Outer*) in which a trainer modeled correct and incorrect applications of the cardinal-count transition and counting out a specified number. A participants indicated whether a muppet had counted out items correctly or not. Explicit feedback for correct modeling included, for instance, "Yes, Cookie Monster you were supposed to count out six cookies, and you stopped counting out at 'six' cookies." Explicit feedback for modeling the common error of counting out all 10 items available was, for example, "Wrong Cookie Monster; you were supposed to count out *six* cookies, but you did not stop counting at *'six.'* Next time stop counting when get to the number you're asked to give."

Practice counting out involved the same instructional games but were implemented differently to avoid providing Teach-to-Target participants Level-2 instruction. For example, both interventions involved the practice game *Animal Spots.* To determine total number of spots (pegs) for a leopard, HLP children drew a dot card, counted the dots on it to determine the total (applied the CP), and then took the number of pegs counted (applied the cardinal-count transition). Children in the Teach-to-Target intervention drew a number card, read (or was helped to read) the written number on the card, and retrieved number of pegs read. See Appendix C in the Supplemental materials for details regarding the interventions.

The five measures to ensure fidelity are delineated in Exhibit C.5 of Appendix C in the Supplementary materials. In terms of implementation fidelity, the HLP participants completed 94 % of the 48 Level-1 (Session 1 & 2) prescribed activities, 99 % of the 159 Level-2 and -3 (Session 3–8) activities, and 100 % of the 152 Level-4 (Session 9–12) activities. The Teach-to-Target participants completed 100 %, 95 %, and 100 % of their Level-1, Session 3–8, and Level-4 activities.

### 5.4. Design and procedures

Eligible children were randomly assigned to the HLP intervention or Teach-to-Target condition. The Teach-to-Target intervention served as an active control (a comparison group that receives a treatment different from that of the experimental condition) and was implemented simultaneously with the HLP intervention. Random assignment and an active control served to counter various threats to internal validity, including a history effect, novelty effect, and dosage of cardinality training overall and the target cardinality training particularly.

All testing and training were individually administered by one of two project staff (the senior author and an assistant) in one of two project-designated rooms within a short walking distance from a child's classroom. Both trainers were keenly aware of treating participants in both conditions equally and consciously tried to deliver each with fidelity and—as positive assent had to be secured for every session—with an enthusiastic and a positive manner. To further ensure the high quality of both interventions, every session was observed by one of two other assistants who recorded fidelity and intervened should a question or probe be missed. The timing of testing and training, as well the tasks used for the pretest and posttest are summarized in Table 3. Post-testing was done two weeks after the completion of training to assess retention of target-level learning.

### 5.5. Data analyses

As the assumptions of a parametric test were not met, a non-parametric test (Mann-Whitney *U* Test) was used to compare the pretest and posttest scores within a group and posttest results of the two conditions on the dependent variables. As the pretest scores on the two dependent variables were 0 or nearly so, the two groups had equivalent baselines for these measures. Qualitative analyses of strategies and errors supplemented the quantitative analyses.

## 6. Results and discussion

### 6.1. Pretest and session-1 and -2 training performance

Participants from both conditions exhibited similar developmental competencies on the pretest. One HLP child could verbally count to four; one Teach-to-Target child could count to six and another to nine. All other children could count to ten. Two Teach-to-Target participants used "two" to indicate "many," and three HLP children could subitize to four. All other children were 2-knowers. Similarly, the training logs revealed that only one HLP child (50 % correct) and one Teach-to-Target child (64 % correct) failed to correctly subitize 1 on at least two-thirds of the opportunities in Training Sessions 1 and 2. One HLP and two Teach-to-Target children qualified as 1-knowers. Three HLP and two Teach-to-Target participants met the criterion for 2-knower, and another Teach-to-Target child nearly did so (60 % correct on trials of 2). Two HLP children and one Teach-to-Target child qualified as a 3-knower.

Although the Teach-to-Target participants had a somewhat higher one-to-one counting accuracy rate, all children appeared to know and generally apply the one-to-one counting principle (e.g., five HLP and five Teach-to-Target children were accurate on at least 4.5 of 6 trials). As Table 4 shows, participants initially exhibited little or no success on any of the other measures, including Task 6 (give 5–7) and Task 7 (stop at *n*), which subsequently served as the dependent measures. As expected, the HLP and Teach-to-Target groups did not differ significantly on any pretest measure, and participants in both groups were ready to achieve meaningful 1-to-1 counting (Level 2).

### 6.2. Delayed posttest and session-9 to -12 training performance

Posttest performance on the counting-out or give-*n* task (Task 6) and the stop-at-*n* task (Task 7) were strongly correlated ($r[12] = .81$, $p < .001$), supporting the validity of the latter. Among children who were unsuccessful on all stop-at-*n* trials ($n = 8$) or who were correct on only a single make-up trial ($n = 1$), seven were unsuccessful on all give-*n* trials. All five children who were successful on half or more of the stop-at-*n* trials were correct on half or more of the give-*n* trials. These results indicate that procedural fluency and conceptual understanding are closely related and that non-success on the counting-out (large-*n* give-*n*) task is typically linked to a conceptual issue, not merely a performance failure.

As shown in Table 4, the HLP intervention had a significantly and (as indicated by large effect sizes) substantially greater impact than the Teach-to-Target instruction on the procedural fluency (Task 6) and conceptual understanding (Task 7). Although both groups received similar Level-4 instruction and the essentially the same dosage of target instruction, only the HLP group—which first received sustained remedial training on subitizing to four (Level 1), the hypothesized developmental prerequisite CP (Level 2), and CP applications (Level 3)—generally benefitted in a substantive manner. Six of seven HLP, but none of seven of the Teach-to-Target, participants were successful on half or more of the give-5-to-7 trials (Task 6), which is hypothesized by Fuson (1988) to require understanding of the cardinal-count transition ($p = .005$, one-tailed, Fisher Exact test). Five of seven HLP participants but only one of seven Teach-to-Target children were successful on half or more of the stop-*n* task, which directly measures the cardinal-count transition. Though not significant ($p = .103$, one-tailed, Fisher Exact test) due to low power, the difference yields a strong effect size (Odds Ratio = 15; Yule's Q = + 0.875).

As Table 5 shows, the quality of HLP participants' responses on the procedural-fluency dependent measure (give-*n* task) was clearly superior to that of Teach-to-Target children. The former accurately counted out the requested number nearly half the time, whereas Teach-to-Target children did so none of the time.

The classic end-with-*n* error (counting out all the items but labeling or relabeling the last item with the requested number) was the most frequent error made by the HLP participants but somewhat less prevalent among Teach-to-Target children. For example, asked to give Cookie Monster six cookies, one child counted out all 10 available but labeled the tenth cookie "six," not "ten." The poorest performing child from the HLP group accounted for three-fourths of such errors by the group; two children accounted for 86 % of such errors made by the Teach-to-Target group. Such end-with-*n* error is consistent with the children learning that the requested number is important and should be remembered but not with understanding the cardinal-count transition: The requested number specifies when to stop the counting-out process. The end-with-*n* error, then, is analogous to the last-word rule in that both reflect partial learning of a cardinality concept by rote and a transition to a full understanding of it.

The HLP participants made relatively few of the four most serious counting-out errors (the last four columns of Table 5), whereas Teach-to-Target participants frequently did so. Simply counting out all the items was relatively rare in both groups. After doing so, one Teach-to-Target child noted, "That's five," indicating she remembered the requested amount but did not understand its implications for her counting effort. For the other most serious errors—major counting-out errors, putting out (e.g., simply grabbing) a subset of items, and putting out all the items—the rate for Teach-to-Target participant was much greater than that for the HLP children. Indeed, the grabbing and putting-out-all errors accounted for more than half the responses of the former. For three Teach-to-Target children (but no HLP participants), putting-out errors accounted for nearly all their responses. This is consistent with these children not benefitting from the modeling and practice of the counting-out procedure because they did not understand the conceptual basis for the procedure (the cardinal-count transition).

As Table 6 shows, the quality of HLP participants' responses on the conceptual dependent measure (stop-at-*n* task) was also clearly superior to that of Teach-to-Target children. The former accurately stopped a count at the requested number almost half of the time, whereas the latter never did. Furthermore, HLP participants were off by 1 on almost a fifth of all their responses, compared to Teach-to-Target children's a tenth of a time. Off-by-1 errors may have been due to a competence or performance failure. Regarding the latter, some children in their excitement and anticipation to help the Muppet may have precipitously and accidentally stopped the count one number before reaching the target number. Other children may have been slow in stopping the count at the requested amount, because applying the newly learned cardinal-count concept was not yet automatic. In contrast, Teach-to-Target participants often responded cluelessly and made more than twice as many of highly inaccurate errors as did HLP participants. Specifically, three of the seven Teach-to-Target children responded haphazardly on all trials, two did so on most trials, and two never stopped the count at the requested number on most trials.

Performance on the dependent measures paralleled performance on a (Session-7 to -11) training task designed to foster Level-2 understanding and supports the hypothesis that the CP is a prerequisite for a cardinal-count transition and successful counting-out performance on the give-*n* task. The one HLP participant who was unsuccessful on both the conceptual (stop-at-*n*) task and the procedural-fluency (give-5-to-7) task was successful on only 10% of the how-many questions of the *Hidden-Chips* training task. The negligible success on this training task indicated she did not learn the CP (or even a last-word rule). A second HLP participant who had modest success on both dependent measures had struggled on this CP training task (i.e., correctly responded to only 50% of the how-many questions overall). The remaining five HLP children, including two with a strong performance on one dependent measure and two with a strong performance on both measures, successfully answered the how-many questions at least two-thirds of the time overall during the training. Over the last three sessions, one of these five was correct 67%; one, 80%; and three, 100% of the time.

## 7. Conclusions

### 7.1. The value of HLP-based (cardinality) instruction

Like Clements et al. (2019, 2020, 2021), the present results support Assumption #1—progressively teaching one level above a child's existing level on an HLP is more efficacious than skipping levels and directly teaching the target level. The present results are the first to extend the applicability of this assumption to the domain of early verbal-based cardinality development. Although all participants in both conditions began the study at a similar developmental level, the similar Level-4 target training had a positive impact on all but one of the seven HLP participants and none of the seven Teach-to-Target participants. The key difference was that the HLP participants received training on the hypothesized precursors and conceptual prerequisite for the target-level instruction (the counting-out procedure and its underlying rationale the cardinal-count transition) but the Teach-to-Target participants did not.

The present results support reform efforts that encourage using formative assessment in conjunction with a HLP in several ways, especially in cases where an earlier level is a necessary condition for a later one (Baroody & Pellegrino, 2023). One is the need to identify a student's current developmental level and teach one level beyond it so that planned or targeted instruction is within a child's zone of proximal development. Another is to ensure that, if planned or targeted instruction is several levels beyond a child's current level, the intermediate developmental levels—particularly concepts *necessary* to understand concepts and procedures at later levels—are identified and consolidated before beginning the target instruction.

### 7.2. The count-cardinal concept (CP) as necessary for the cardinal-count concept

Consistent with other research that support Fuson's hypothesis (e.g., Baroody & Lai, 2022; Fuson, 1988), the present results fit (almost) perfectly the expected outcome if the CP (a conceptual basis for Level 2) is a necessary condition for fluently counting-out a requested number of items and its conceptual basis, namely the cardinal-count transition (Level 4). The six HLP children who benefitted from the Level-2 (CP) training had at least some success on the procedural-fluency measure, and five were successful on half or more of the conceptual trials. Moreover, when wrong, these participants tended to make minor errors indicative of a performance failure due to task demands rather than a lack of conceptual understanding. Put differently, all but one HLP participant was well on the way to at least approximating (a) procedural fluency with counting out and (b) understanding its conceptual basis.

In stark contrast, the HLP child who apparently did not construct the requisite Level-2 knowledge and all seven Teach-to-Target participants who did not have CP training failed to achieve even a modest approximation of counting-out fluency—as evidenced by their almost complete lack of success on the give-*n* task, negligible proportion of close errors, and considerable frequency of major errors. Such ineffective responses are indicative of a lack of conceptual understanding—a conclusion supported by an almost complete lack of success on the stop-at-*n* task (a direct measure of cardinal-count transition). Unlike the six HLP children who made progress, these eight participants benefitted negligibly, if at all, from modeling the counting-out procedure and its conceptual rationale (e.g., remember the requested number and stop counting out when it is reached), because they did not have a basis for assimilating this Level-4 instruction. In brief, without understanding the CP, learning its inverse (the cardinal-count transition) and the procedure it informs was beyond their reach.

### 7.3. Limitations and future directions

It could be argued that the HLP condition in present research does not fully embody an HLP-instructional approach as envisioned in Realistic Mathematics Education (Streefland, 1991) and the Building Blocks curricula. For example, HLP children were *not* (as normally done with HLP-based instruction) tested for mastery of a level before proceeding onto the next level. However, the focus of the research (rigorously evaluating the efficacy and an assumption of HLP-based instruction) required a controlled set of teaching strategies to eliminate alternative hypotheses such as dosage. Even so, the diluted HLP-based instruction provided was clearly superior to the Teach-to-Target intervention.

It could be questioned whether research about the value and assumptions of HLP-based instruction is even needed, because the results seem obvious. However, the present study is only the fifth of our 10 experiments that evaluated the efficacy of HLP-based instruction and the first of four that examined cardinality development to achieve statistical significance (see Baroody et al., 2023, for a review). Aside from the methodological considerations required and the difficulty of achieving treatment fidelity, a hard-learned lesson of our research project on HLP-based instruction is that the value of such an approach for different topics and whether levels (and how many) can be skipped is not obvious. It depends on such factors as: (a) the nature of the relation between successive levels, (b) the degree to which successive levels differ, and (c) the number of paths to target knowledge. Following an HLP is usually essential when it entails *necessary* prerequisites. However, it may be only highly to mildly helpful when it entails facilitative precursors and—as Lesh and Yoon (2004) noted—not particularly helpful when knowledge is constructed as a web. Even if an HLP involves a level necessary for a subsequent level, it does not preclude skipping a level if the latter level is not much of a conceptual leap from the former, if there is an alternative path to the target that merely involves facilitative relations, or if an instructional intervention has the efficacy to teach two levels simultaneously.

A possible limitation of the present research was that the trainers and testers knew the purpose of the research and assignment of the participants. Thus, unconscious bias cannot be discounted. Another limitation is that the original plan to work with a sample twice the size of that of the present study was interrupted by the State's stay-at-home order due to the COVID-19 pandemic. So, it could be argued that generalizing or otherwise drawing conclusions from only seven children who were exposed to the experimental treatment and seven who were not cannot be justified. The small sample is not a threat, though, to internal validity—a plausible alternative explanation for the conclusion that teaching successive levels of the cardinality HLP is more efficacious than teaching Level 4 by direct transmission. Indeed, despite a relatively small sample of 14, the underpowered experiment produced a statistically significant difference favoring the HLP intervention on both dependent measures.

Slavin and Smith (2009) caution about drawing inferences from even the large effect sizes of small-*n* studies, because such studies produce less reliable and replicable estimates of program impact than do large-*n* studies. They further note that the most important source of this greater variability may be what Cronbach et al. (1980) call "superrealization"—high implementation fidelity due to better monitoring and more input by experimenters than would be available at scale. Moreover, the sample was drawn from a public preschool for children potentially at-risk defined by the school's admission criteria as eligible for free or reduced lunch, medical reasons (e.g., low birth weight), or ethnic background. Thus, the present study has limited external validity and needs to be replicated with a larger, more representative sample.

The present research was designed to test Assumption 1 and Fuson's (1988) hypothesis directly and rigorously but not Sarnecka and Carey's (2008) single-step, single-concept hypothesis or Brousseau's (1997) perspective. Sarnecka and Carey could argue that the HLP participants significantly and substantially outperformed the Teach-to-Target children because the former received training on the CP, which is the basis for both meaningful one-to-one counting and counting out, and the latter did not. Moreover, on their view, the target training on the irrelevant cardinal-count transition would not be expected to help either group.

Left open, then, are two questions:

- Which is more efficacious for fostering fluency with counting out collections beyond the subitizing range, CP training alone (Sarnecka and Carey's view) or CP training and then training involving the cardinal-count transition and counting-out procedure (Fuson's, 1988, view)?
- How do children go from knowing the CP, which applies to counting a collection to determine its a total (situations involving set-to-number mapping) to knowing when to stop the counting-out process (situations involving number-to-set mapping)?

The latter question is not addressed by Sarnecka and Carey's theorizing or research. Interestingly, many children who know the CP respond to the give-*n* task by counting out a collection, then counting the collection to check, and finally adjusting their response up or down (Barbara W. Sarnecka, personal communication, April 1, 2024). Counting out an approximate number of items, counting the results, and adding to or taking away from the result to match the cardinal number requested may be the only strategy open to children who understand the CP but *not* the cardinal-count transition and, thus, have no way of knowing when to stop the counting-out process. Reflecting on how to circumvent such a cumbersome strategy may prompt children to consider where exactly in the counting-out process they should stop to create a collection that if counted would result in the requested number (i.e., induce the cardinal-count transition). After all, it would make sense that counting a collection to determine its total (a set-to-number mapping) would require one concept (the CP) and that counting out a specified number (a number-to-set mapping) would require the inverse concept (cardinal-count transition). This account also explains why children fluent in creating collections typically do not count the collection to check the counting-out process.

It could be argued that, from Brousseau's (1997) perspective, an alternative explanation for the results was the HLP training involved instruction that bridged counting a collection, counting-out an equivalent collection with different items, *and* establishing a one-to-one correspondence between the two collections, and the Teach-to-Target training did not. For example, the HLP version of the *Animal Spots* game involved counting a collection of dots, and then counting out an equivalent collection of pegs but Teach-to-Target version did not. HLP, but not Teach-to-Target, participants then had an opportunity to mentally create a one-to-one link between the dots and pegs or—using the pegs as a bridge—between the dots and holes in an animal board. However, it seems unlikely that preschoolers would be able to visualize or imagine a one-to-one correspondence between the previously counted dots and the counted-out pegs, let alone between the dots and holes. Without one-to-one correspondence to check the validity of counting-out efforts, Brousseau's perspective predicts that neither the Teach-to-Target nor the HLP target training should have been successful. The success of HLP training indicate that establishing a one-to-one correspondence between a counted collection and one that is counted out is not a necessary condition for learning the cardinal-count transition and the skill of counting-out a collection but that the CP and accurate one-to-one counting are. In brief, the present results are consistent with Fuson's (1988) sequential hypothesis but inconsistent with Brousseau's simultaneous-teaching recommendation.

Even so, for children who have already constructed the CP, Brousseau (1997) view raises the interesting instructional possibility that establishing a one-to-one correspondence between a counted collection and one that is counted out might *facilitate* the construction of the cardinal-count transition and fluency with counting-out. For example, a teacher could have a child could count *n* dots on a card (How many dots?"), count out *n* items ("Okay, take six pegs"), and finally check whether the latter is correct via matching (placing the pegs on the dots on the card). Whether such a teaching approach is significantly more efficacious than the HLP intervention used in the present study needs to be evaluated.

### 7.4. Summary

A novel variation of an experimental design was employed to test a key assumption for using hypothetical learning progressions (HLP) to improve mathematics instruction: Teaching each successive level of a progression is more efficacious than skipping lower levels and teaching the target level directly. The present evaluation of this assumption entailed teaching different aspects of cardinality knowledge to HLP and Teach-to-Target participants, namely the prerequisite cardinality knowledge for understanding and implementing the target knowledge of counting-out a specified number of items and the unrelated cardinality knowledge of reading written numbers, respectively. Both groups received basically the same target instruction, namely explicit instruction on the conceptual basis and procedure for counting-out. Quantitative and qualitative analyses of the results clearly indicated teaching the same thing in the same way did not produce the same result when prior instruction differs–when (necessary) prerequisite knowledge is considered and ensured (as in the HLP approach) or not (as in the Teach-to-Target approach). Specifically, adhering to Piaget's principle of assimilation (the HLP approach) resulted in an appreciable understanding of the conceptual rationale for counting-out and procedural fluency with the skill but violating the principle (the Teach-to-Target approach) did not. Adherence to the principle entailed assessing what level a child was at on the cardinality HLP and fostering the intermediate levels in the progression before attempting to teach the target level. In brief, the pattern of results supports the logic of using HLPs and formative assessment and does so in a new way.

## CRediT authorship contribution statement

**Arthur James Baroody:** Writing – review& editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Douglas H. Clements:** Writing – review & editing, Funding acquisition, Conceptualization. **Julie Sarama:** Writing – review & editing, Funding acquisition, Conceptualization.

## Author contribution

The first author was the primary contributor to the study's conception and design. Material preparation, data collection, and analysis were supervised by the first author. The first author was the primary writer. The second and third authors reviewed and provided feedback on all aspects just mentioned. The authors collaborated on the final version of manuscript, and all authors read and approved it.

## Consent to participate

Informed parental consent was obtained before testing, and positive assent was obtained from a child for each testing and training session.

## Consent for publication

Not applicable.

## Ethics approval

The protocol for the present research was approved by the University of Illinois IRB.

## Funding

The preparation of this report was supported by the Institute of Education Science Grant R305A150243 to the authors and the National Science Foundation Grant nos. 1621470 and 2201939 to the first author. The opinions expressed are solely those of the authors and do not necessarily reflect the position, policy, or endorsement of the Institute of Education Science or the National Science Foundation.

## Conflict of interest

The authors have no competing interests to declare that are relevant to the content of this article.

## Supplementary Information

The online version contains Supplementary material available at.

## Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Arthur J. Baroody reports financial support was provided by US Department of Education. Douglas H. Clements reports financial support was provided by US Department of Education. Julie Sarama reports financial support was provided by US Department of Education. Arthur J. Baroody reports financial support was provided by National Science Foundation. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data will be made available on request.

## Appendix A.Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jmathb.2024.101178.

## References


Balfanz, R. (1999). Why do we teach children so little mathematics? Some historical considerations. In J. V. Copley (Ed.), *Mathematics in the early years* (pp. 3–10). National Council of Teachers of Mathematics.

Baroody, A. J. (2016a). Curricular approaches to introducing subtraction and fostering fluency with basic differences in grade 1. In R. Bracho (Ed.), *Monograph of the Journal of Pensamiento Numérico y Algebraico (Numerical and Algebraic Thought): 10(3). The development of number sense: From theory to practice.* University of Granada. https://doi.org/10.1080/17405629.2016.1147345.

Baroody, A. J. (2016b). Using number and arithmetic instruction as abasis for fostering mathematical reasoning. In M. T. Battista (Ed.), *Reasoning and sense making in the mathematics classroom: Pre-K—Grade 2* (pp. 22–69). National Council of Teachers of Mathematics.

Baroody, A. J., & Lai, M.-L. (2022). The development and assessment of counting-based cardinal-number concepts. *Educational Studies in Mathematics, 111*, 185–205. https://doi.org/10.1007/s10649-022-10153-5

Baroody, A. J., Mix, K., Kartal, G., & Lai, M. L. (2023). The development and assessment of early cardinal-number concepts. *Journal of Numerical Cognition, 9*(1), Article 10.5964/jnc.10035.

Baroody, A. J., & Pellegrino, J. W. (2023). Assessment for learning. In C. J. Harris, E. N. Wiebe, S. Grover, & J. W. Pellegrino (Eds.), *Classroom-based STEMAssessment: Contemporary issues and perspectives. Community for AdvancingDiscovery Research in Education (CADRE). Education Development Center, Inc.*.

Benoit, L., Lehalle, H., & Jouen, F. (2004). Do young children acquire number words through subitizing or counting? *Cognitive Development, 19*, 291–307. https://doi.org/10.1016/j.cogdev.2004.03.005

Bereiter, C. (1986). Does direct instruction cause delinquency? Response to Schweinhart and Weikart. *Educational Leadership, 44*, 20–21.

Brousseau, G. (1997). In N. Balacheff, M. Cooper, R. Sutherland, V. Warfield, & Trans.) (Eds.), *The theory of didactical situations in mathematics*. Dordrecht: Kluwer.

Clark, R. C., Nguyen, F., & Sweller, J. (2006). Efficiency in learning: Evidence-based guidelines to manage cognitive load. *Pfeiffer*.

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.

Clements, D. H., & Sarama, J. (2021). Learning and teaching early math: The learning trajectories approach. In *Routledge*.

Clements, D. H., Sarama, J., Baroody, A. J., & Joswick, C. (2020). Efficacy of a learning trajectory approach compared to a teach-to-target approach for addition and subtraction. *ZDM Mathematics Education, 52*(4), 637–649. https://doi.org/10.1007/s11858-019-01122-z

Clements, D. H., Sarama, J., Baroody, A. J., Joswick, C., & Wolfe, C. (2019). Evaluating the efficacy of a learning trajectory for early shape composition. *American Educational Research Journal, 56*(6), 2509–2530. https://doi.org/10.3102/0002831219842788

Clements, D. H., Sarama, J., Baroody, A. J., Kutaka, T. S., Chernyavskiy, P., Joswick, C., Cong, M., & Joseph, E. (2021). Comparing the efficacy of early arithmetic instruction based on a learning trajectory and teaching-to-a-target. *Journal of Educational Psychology*. https://doi.org/10.1037/edu0000633

Council of Chief State School Officers . (2010). *Common core state standards: Preparing America's students for college and career*. Retrieved from 〈http://www.corestandards.org/〉.

Cronbach, L. J., Ambron, S. R., Dornbusch, S. M., Hess, R. O., Hornik, R. C., & Phillips, D. C. (1980). *Toward reform of program evaluation: Aims, methods, and institutional arrangements*. Jossey-Bass.

Dorier, J.-L. (2015). Key issues for teaching numbers within Brousseau's theory of didactical situations. *The twenty-third ICMI study: Primary mathematics study on whole numbers*. China: Macao.

Engel, M., Claessens, A., & Finch, M. A. (2013). Teaching students what they already know? The (mis)alignment between mathematics instructional content and student knowledge in kindergarten. *Educational Evaluation and Policy Analysis, 35*, 157–178. https://doi.org/10.3102/0162373712461850

Ferguson, R. F. (1998). Can schools narrow the Black-White test score gap? In C. Jencks, & M. Phillips (Eds.), *The Black-White test score gap* (pp. 318–374). Brookings Institution Press ([ISBN: 9780815746102]).

Frye, D., Baroody, A. J., Burchinal, M. R., Carver, S., Jordan, N. C., & McDowell, J. (2013). *Teaching math to young children*: *A practice guide.* Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE). 〈http://ies.ed.gove/ncee/wwc/practiceguide.aspx?sid=18〉.

Fuson, K. C. (1988). *Children's counting and concepts of number*. Springer-Verlag.

Geary, D. C., vanMarle, K., Chu, F. W., Rouder, J., Hoard, M. K., & Nugent, L. (2018). Early conceptual understanding of cardinality predicts superior school-entry number-system knowledge. *Psychological Science, 29*(2), 191–205. https://doi.org/10.1177/0956797617729817

Gelman, R. (1993). A rational-constructivist account of early learning about numbers and objects. In D. L. Medin (Ed.), *The Psychology of Learning and MotivationAdvances in research theory* (pp. 61–96). Academic Press.

Institute of Medicine (IOM) and National Research Council (NRC). (2015). *Transforming the workforce for children birth through age 8: A unifying foundation*. National Academy Press (〈www.nap.edu/catalog/19401/transforming-the-workforce-for-children-birth-through-age-8-a〉).

James, W. (1958). *Talks to teachers on psychology: And to students on some of life's ideals*. NY: W. W. Norton & Company.

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

Kilday, C. R., Kinzie, M. B., Mashburn, A. J., & Whittaker, J. V. (2012). Accuracy of teachers' judgments of preschoolers' math skills. *Journal of Psychoeducational Assessment, 30*(2), 48–158. https://doi.org/10.1016/j.ecresq.2014.06.007

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*, 75–86.

Kilpatrick, J. , Swafford, J. , Findell, B. , (2001). *Adding it up: Helping children learn mathematics. National Academy Press*

Le Corre, M., Van de Walle, G. A., Brannon, E., & Carey, S. (2006). Revisiting the performance/competence debate in the acquisition of counting as a representation of the positive integers. *Cognitive Psychology, 52*, 130–169. https://dx.oi.org/10.1016/j.cogpsych.2005.07.002.

Lehrer, R., & Schauble, L. (2015). Learning progressions: The whole world is NOT a stage. *Science Education, 99*(3), 432–437.

Lee, J. S., & Ginsburg, H. P. (2007). Preschool teachers' beliefs about appropriate early literacy and mathematics education for low- and middle-SES children. *Early Education & Development, 18*(1), 111–143.

Lesh, R., & Yoon, C. (2004). Evolving communities of mind—Where development involves several interacting and simultaneously developing strands. *Mathematical Thinking and Learning, 6*, 205–226.

Li, X., Chi, L., DeBey, M., & Baroody, A. J. (2015). A study of early childhood mathematics teaching in the U.S. and China. *Early Education and Development, 26*, 37–41. https://doi.org/10.1080/10409289.2015.994464

Lobato, J., & Walters, C. D. (2017). A taxonomy of approaches to learning trajectories and progressions. In J. Cai (Ed.), *Compendium for research in mathematics education* (pp. 74–101). National Council of Teachers of Mathematics.
Maloney, A. P., Confrey, J., & Nguyen, K. H. (Eds.). (2014). *Learning over time: Learning trajectories in mathematics education*. Information Age Publishing.
Mou, Y., Zhang, B., Piazza, M., & Hyde, D. C. (2021). Comparing set-to-number and number-to-set measures of cardinal number knowledge in preschool children using latent variable modeling. *Early Childhood Research Quarterly, 54*, 125–135. https://doi.org/10.1016/j.ecresq.2020.05.016
O'Rear, C. D., Kirkland, P. K., & Purpura, D. J. (2024). The how many and give-N tasks: Conceptually distinct measures of the cardinality principle. *Early Childhood Research Quarterly, 66*(1)), 61–74. https://doi.org/10.1016/j.ecresq.2023.08.010
Paliwal, V., & Baroody, A. J. (2020). Cardinality principle understanding: The role of focusing on thesubitizing ability. *ZDM Mathematics Education, 52*(4). https://doi.org/10.1007/s11858-020-01150-0
Piaget, J. (1964). Development and learning. In R. E. Ripple, & V. N. Rockcastle (Eds.), *Piaget rediscovered* (pp. 7–20). Cornell University Press.
Renkl, A. (2005). The worked-out examples principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning*. Cambridge: Cambridge University Press.
Resnick, L. B., & Ford, W. W. (1981). *The psychology of mathematics for instruction*. Erlbaum.
Rosenshine, B. (2009). The empirical support for direct instruction. In S. Tobias, & T. M. Duffy (Eds.), *Constructivist theory applied to instruction: Success or failure?* (pp. 201–220). Taylor & Francis.
Rosenshine, B. (2012). Principles of instruction: Research-based strategies that all teachers should know. *American Educator, 36*(1), 12–19.
Sarama, J. , & Clements, D.H. (2009). *Early childhood mathematics education research:Learning trajectories for young children*. Routledge.
Sarnecka, B. W., & Carey, S. (2008). How counting represents number: What children must learn and when they learn it. *Cognition, 108*, 662–674. https://doi.org/10.1016/j.cognition.2008.05.007
Schaeffer, B., Eggleston, V. H., & Scott, J. L. (1974). Number development in young children. *Cognitive Psychology, 6*, 357–379.
Shavelson, R. J., & Karplus, A. (2012). Reflections on learning progressions. In A. C. Alonzo, & A. W. Gotwals (Eds.), *Learning progressions in science: Current challenges and future directions* (pp. 13–26).
Schwonke, R., Renkl, A., Krieg, C., Wittwer, J., Alven, V., & Salden, R. (2009). The worked-example effect: Not an artefact of lousy control conditions. *Computers in Human Behavior, 25*, 258–266.
Simon, M. A. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for Research in Mathematics Education, 26*(2), 114–145. https://doi.org/10.2307/749205
Slavin, R., & Smith, D. (2009). The relationship between sample sizes and effect sizes in systematic reviews in education. *Education Evaluation and Policy Analysis, 31*(4), 500–506. https://doi.org/10.3102/0162373709352369
Streefland, L. (Ed.). (1991). *Realistic mathematics education in primary school*. Freudenthal Institute, Utrecht University.
Von Glasersfeld, E. (1982). Subitizing: The role of figural patterns in the development of numerical concepts. *Archives De lelott Psychologie, 50*, 191–218.
Vygotsky, L.S. , (1934/1986). *Thought and language* (E. Hanfmann & G. Vakar, Eds. &Trans.). MIT Press.
Wu, H.-H. (2011). *Understanding numbers in elementary school mathematics*. American Mathematical Society.