



The development and assessment of counting-based cardinal number concepts

Arthur J. Baroody¹ · Menglung Lai²

Accepted: 8 April 2022 / Published online: 13 May 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract

The *give- n* task is widely used in developmental psychology to indicate young children’s knowledge or use of the cardinality principle (CP): the last number word used in the counting process indicates the total number of items in a collection. Fuson (1988) distinguished between the CP, which she called the count-cardinal concept, and the cardinal-count concept, which she argued is a more advanced cardinality concept that underlies the counting-out process required by the *give- n* task with larger numbers. One aim of the present research was to evaluate Fuson’s disputed hypothesis that these two cardinality concepts are distinct and that the count-cardinal concept serves as a developmental prerequisite for constructing the cardinal-count concept. Consistent with Fuson’s hypothesis, the present study with twenty-four 3- and 4-year-olds revealed that success on a battery of tests assessing understanding of the count-cardinal concept was significantly and substantially better than that on the *give- n* task, which she presumed assessed the cardinal-count concept. Specifically, the results indicated that understanding the count-cardinal concept is a necessary condition for understanding the cardinal-count concept. The key methodological implication is that the widely used *give- n* task may significantly underestimate children’s understanding of the CP or count-cardinal concept. The results were also consistent with a second aim, which was to confirm that number constancy concepts develop after the count-cardinal concept but before the cardinal-count concept.

Keywords Assessment · Cardinality principle · Counting concepts · Early childhood · *Give- n* task · Learning progression

The development of verbal-based cardinality knowledge—recognizing that a number word such as “three” represents the total number of items in a set (the set’s cardinal value), generally begins at about 2.5 years of age. Subitizing-based small- n recognition—the ability to discern without counting the total number of items in a collection and associate it with an

✉ Arthur J. Baroody
baroody@illinois.edu

Menglung Lai
menglung@gmail.com

¹ College of Education, University of Illinois, Champaign, IL 61853, USA

² Department of Early Childhood Education, National Chiayi University, Chiayi, Taiwan

Table 1 A hypothetical learning progression of key aspects of counting-based cardinal number knowledge

Aspect of cardinal number	Conceptual basis	Mapping	Sample measure
<i>First (pre-counting or subset-knowers) phase of cardinality development—before meaningful 1-to-1 counting (before understanding of the CP)</i>			
Subitizing-based (small- <i>n</i>) number recognition—often called <i>n</i> -knower levels	<i>Cardinal representation/concept</i> of a small number underlies the reliable ability to subitize (immediately recognize 1, 2, or 3)	Set-to-word (via verbal subitizing)	How-many task with 1 to 4 items without counting
Subitizing-based (small- <i>n</i>) set creation—also often called <i>n</i> -knower levels	<i>Cardinal representation/concept</i> of small numbers can be used to subitize when 1, 2, or 3 items have been put out (i.e., reliably stop the set-creation process)	Word-to-set (via verbal subitizing)	Give- <i>n</i> task with 1 to 4 items without counting
<i>Second (counting and generalized number) phase of cardinality development—meaningful 1-to-1 counting (i.e., with understanding of the CP)</i>			
Possible level 1 (PL1): meaningful 1-to-1 counting	<i>Count-cardinal concept</i> or CP (last number word = total)	Set-to-word	How-many task with counting (tasks 1 & 3)
PL2: applications of CP: <i>number constancy concepts</i>	2A. conservation of cardinal identity: addition or subtraction but not irrelevant physical transformations, changes total (cardinal number) of a counted set 2B. Cardinal equivalence: counted sets with the same last number word (cardinal number) are equal in number despite looking different	Set-to-word-to-transformed set Comparing two set-to-word mappings	Transform-set task (task 4) Compare-set task (task 5)
PL3: set creation	<i>Cardinal-count concept</i> (a cardinal number = the last number word would be if a set was counted)	Word-to-set	Give- <i>n</i> task with counting (task 2)

CP, cardinality principle. Bold text in column 4 (sample measure) indicates a task used in the present study.

appropriate number word (Kaufman et al., 1949)—is typically the children’s initial means of verbally enumerating a set’s total (see Table 1). For example, the ability to subitize two entails immediately recognizing that sets such ●●, ■, or ↗ are all examples of pairs or a unit and another unit (concept of two) and should be labeled “two,” but a single item (e.g., ●, ■, or ↗) or more than a pair of items (e.g., ●●●, ■■, or ↗ ↗ ↗) are “not two.” There is wide agreement that the first phase unfolds in a stepwise manner in the order of magnitude—commonly called the *n*-knower levels (Condry & Spelke, 2008; Le Corre & Carey, 2007, 2008; Le Corre et al., 2006; Sarnecka & Carey, 2008; Wynn, 1990, 1992). A 1-knower can reliably recognize and label single items as “one” or can give one item upon request but cannot do so for larger numbers; a “2-knower” can reliably recognize and give sets of one and two but not larger numbers; and so forth up to the 4-knower level (but see Sella et al, 2021).

There is broad agreement that the first phase of verbal-based cardinality development provides a basis for the second phase based on 1-to-1 counting, which greatly extend the children’s number competence (Baroody et al., 2006; Baroody & Purpura, 2017; Benoit et al., 2004; Carey & Barner, 2019; Fischer, 1992; Klahr & Wallace, 1976; von Glasersfeld, 1982; but cf. Gallistel & Gelman, 2000; Nieder, 2017). Some researchers hypothesize that counting-based cardinality knowledge unfolds in a stepwise manner, with an earlier level of understanding provides a foundation for a later level (see Table 1) (Baroody & Purpura, 2017; Frye et al., 2013; Fuson, 1988; but cf. Le Corre et al., 2006; Sarnecka & Carey, 2008). The present research addressed two issues regarding this hypothetical learning progression (HLP) for the second phase of cardinality development.




1 Rationale

1.1 A hypothetical learning progression for the second phase

Possible level 1 (PL1): count-cardinal concept and meaningful one-to-one counting Children may learn to count by rote—without realizing its purpose is to determine how many. In time, they recognize that counting—like subitizing—can be used to identify the total number of items in a collection. Meaningful object counting requires constructing the cardinality principle (CP) or what Fuson (1988) called the count-cardinal concept: the understanding that the last number–word used when counting has special significance because it represents the total number of items in a collection (PL1 in Table 1). (Although Fuson’s term is more precise, the terms CP, CP knowledge, and CP-knower will henceforth be used because of their greater familiarity and brevity.) Construction of the CP enables children to move from understanding a few small numbers (subset-knowers) to understanding numbers in general (e.g., generalizing local insights with small numbers to—in principle—all numbers). The CP may provide a basis for constructing more advanced cardinality concepts, PL2 and PL3.

PL2: counting-related number constancy concepts Conservation of cardinal identity entails recognizing that a counting-generated cardinal label such as “five” is still applicable even if the physical appearance of a collection is changed and that it does change if items are added to or subtracted from the set (PL2A in Table 1). Cardinal equivalence involves recognizing that counting can serve to determine whether two collections are equal in

number, even if they differ in physical appearance. For example, a linear array of 5 squares and a haphazard array of 5 dots are equal in number if both have the same counting-generated cardinal label “five” (PL2B).

PL3: cardinal-count concept and the counting-out procedure Children can use subitizing to produce up to about four requested items. With larger numbers, they must count out items to do so. Initially, children often do not stop the counting-out process at the requested number and simply count all the items provided. To overcome this common no-stop error, Fuson (1988) hypothesized children first need to construct the cardinal-count concept: understanding that a cardinal number would be the last number word *if* a set was counted. Specifically, this concept provides the rationale for monitoring the counting-out process and stopping it at the requested number (PL3). For instance, for “give me *five* chips,” the cardinal-count concept underlies the ability to recognize that “five” is the number word at which the counting-out process should stop. In effect, the cardinal-count concept, which entails a word-to-set mapping, is the inverse of the count-cardinal concept, which involves a set-to-word mapping. Word-to-set mapping involves starting with a number word and its associated general concept and creating a specific example of the number (e.g., relating the number word “five” and a cardinal concept of five such as five units or four and one more to particular cases such as , , or ). Set-to-word mapping entails starting with a specific example of a number and relating it to a number word and its associated general concept.

Both PL1 and PL3 are commonly called the cardinality principle knower (CP-knower) level. To distinguish between these levels, “CP,” “CP knowledge,” or CP-knower” will hereafter refer only to PL1 (count-cardinal) knowledge unless explicitly qualified (e.g., “CP-knower level as gauged by the give-*n* task”).

1.2 Issues addressed by the present study

Issue 1: as hypothesized by Fuson (1988), does the CP develop before and serve as a developmental prerequisite (necessary condition) for a distinct cardinal-count concept? Existing evidence has not clearly established whether these concepts are distinct and emerge sequentially (Fuson) or are one in the same and emerge simultaneously (Sarnecka & Carey, 2008). With collections beyond the subitizing range, research generally indicates that success on the how-many task precedes that on the give-*n* task (but cf. Le Corre et al., 2006). Although this developmental trend is consistent with Fuson’s hypothesis, alternative explanations have been adduced. One is performance factors—the greater task demands of the give-*n* task. For example, counting out a collection requires holding the requested number in working memory and comparing this cardinal number to each number word used to count out items (Frye et al., 1989; Le Corre et al., 2006; Resnick & Ford, 1981).

Fuson (1988) offered that the how-many task may *overestimate* knowledge of the CP, because children may first respond to how-many questions by applying a non-meaningful “last-word rule.” That is, initially, they may simply recognize that repeating the last number word used in the counting process is an acceptable response to a how-many question and only later construct the CP (realize that last number word represents the total).

Indeed, Sarnecka and Carey (2008) found that most 2-knowers to 4-knowers—as well as nearly all CP-knowers (as measured by the give-*n* task)—were successful on the how-many tasks with collections of 5 and 7. They concluded that children “apply [a last-word]

rule long before they demonstrate any understanding of the cardinality principle on tasks such as Give-N” (p. 669). To support their claim that success on the how-many task merely indicates pseudo-knowledge of the CP (i.e., the readily achieved and meaningless last-word rule) and success on the give- n task reflects genuine understanding of the CP, Sarnecka and Carey cited the results of Le Corre et al.’s (2006) Experiment 2, which involved a Counting-Puppet task to measure the CP. The task entailed telling children a puppet wanted n cookies (6, 7, and 8 cookies on trials 1, 2, and 3, respectively); counting out $n - 1$, n , or $n + 1$ (5, 7, and 9 cookies, respectively), and asking, “Is that n (6, 7, and 8, respectively)?” Whereas subset-knowers responded at a chance level on the Counting-Puppet task, CP-knowers (as measured by the give- n task) responded at an above chance level. Sarnecka and Carey (2008, p. 672) concluded:

We are inclined to doubt [that the performance gap between the how-many and give- n tasks we observed is due to the fact the [CP] is learned earlier than the cardinal-count concept], because ... Le Corre et al. (2006) [found] cardinal-principle knowers [as measured by the give- n task] pass and subset-knowers fail [the Counting-Puppet task] ... The main difference is that our How-Many task uses the specific phrase ‘how many’ [which can elicit a correct response via the last-word rule], whereas the Counting Puppet task does not. There is no obvious reason why either task should be a better test of the [CP] than the other [or] why subset-knowers should succeed at our task and fail at Le Corre’s if the [CP] were the issue.

However, there are two problems with Sarnecka and Carey’s (2008) line of reasoning:

1. Although success on the how-many task can be inflated by applications of a last-word rule, some success on the task may represent the application of CP understanding, and construction of this understanding may occur before learning the cardinal-count concept and its attended success on the give- n task. Sarnecka and Carey’s evidence that the performance gap between the how-many and give- n task is due (solely) to the use of the last-word rule on the former is inconclusive, because success on the how-many task does not indicate whether a child used the last-word rule or the CP, and they did not use other means to make this distinction.
2. Contrary to Sarnecka and Carey’s last conclusion, the HLP in Table 1 suggests a plausible alternative explanation for this pattern of results: differences in the conceptual demands of the tasks. Success on the how-many task, which entails set-to-word mapping, depends on knowing the last-word rule *or* the count-cardinal concept (CP). Success on the give- n task *and* the Counting-Puppet task, both of which involve a word-to-set mapping, depends on understanding the developmentally more advanced cardinal-count concept, not the CP as Sarnecka and Carey (2008) presumed. For example, children who did not understand the cardinal-count concept would not understand, for example, “give six” predicts that the counting-out process should stop at “six” and, thus, would have no basis for success on *either* the give- n or Counting-Puppet task. “Subset-knowers” (as indicated by a lack of success on the give- n task with 5 or more), then, could succeed on the how-many task if they knew the last-word rule *or* the CP but fail the Counting-Puppet task because they do not understand the more advanced cardinal-count concept.

Frye et al. (1989) compared performance on the how-many, are-there- n (e.g., shown card with 2, 3, or 4 dots and asking, “Are there three dots here?”), and give- n tasks. They reasoned that success on the how-many task but not the are-there- n task indicated the use

of the last-word rule; success on both tasks, understanding of the CP; and success on the give-*n* task, understanding of the cardinal-count concept. The success rate on the how-many task was significantly better than that on the are-there-*n* task, which was significantly better than that on the give-*n* task. Frye et al. concluded this pattern of performance corroborated Fuson's (1988) hypothesis that a last-word rule develops before the CP, which in turn develops before the cardinal-count concept. However, Frye et al.'s data were not analyzed at the participant level and did not track development over time.

Fuson (1988; Study 8.2.2) combined the results of the how-many and a follow-up task to distinguish between children who merely used a last-word rule and those who possessed a meaningful understanding of the CP. She used a prediction task to directly gauge understanding of the cardinal-count concept. For instance, children were told a collection had six butterfly stickers and asked, "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. The remaining 22 participants, who were last-word responders or had been taught the last-word rule, failed to exhibit an understanding of the cardinal-count concept.

Though Fuson's (1988) evidence is consistent with her hypothesis that the CP develops prior to the cardinal-count concept and that the former may be a basis for the latter, it is also not conclusive. The sample of children who were successful on the CP task was small, and 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 last-word rule). Moreover, as Fuson observed, the wording of the prediction (cardinal-count concept) 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 up 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. Moreover, their analyses ruled out general cognitive or linguistic demands as a source of performance differences. Mou et al. concluded their results are inconsistent with the common assumption that the set-to-word (e.g., how-many) and word-to-set (e.g., give-*n*) tasks gauge interchangeable concepts and are consistent with multiple dimensions of cardinal number knowledge acquisition.

Although Mou et al. (2021) concluded their results were consistent with Fuson's (1988) hypothesis that the cardinal-count concept requires understanding of the CP, they allowed there might be alternative explanations. The research also did not distinguish among accurate responses on the how-many task due to subitizing, a last-word rule, or the count-cardinal concept.

In contrast to the research previously reviewed, the results of Le Corre et al.'s (2006) Experiment 1 appear to support the claim that the CP and cardinal-count concept are not distinct. These researchers used a "What's on the card?" (WOC) task instead of a how-many task to gauge the CP, because the how-many task may be unclear, confusing, or misleading to young children and *underestimate* CP knowledge. Specifically, Le Corre et al. noted the how-many task may be unclear, because children may not realize that it requires both counting and stating the cardinal value of the collection and may simply count a collection and not state the total (i.e., repeat the last number word). As Gelman (1993) observed, the how-many task may be confusing, because children may adhere to the conversational convention that "we are not supposed to repeat what is known" or obvious (p. 80). For example, a child might assume that counting a collection of six items is sufficient for addressing the how-many question. Finally, the task may be misleading, because children might misinterpret a follow-up how-many question to elicit the cardinal value as

indication they counted or responded incorrectly and should count again or change their answer—both of which are scored as incorrect (i.e., not knowing the CP). The WOC task involved presenting a card with a collection of one to seven stickers and asking simply, “What’s on this card?” Le Corre et al. (2006) found that 92% of their 52 participants age 2.1 to 4.0 years were classified as a CP-knower or not on both the WOC task and the give-*n* task. They concluded that the “tasks provided overwhelmingly consistent pictures of what children understand about how counting represents number” (p. 151).

However, the results of Le Corre et al.’s (2006) Experiment 1 are far from clear-cut. 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. These results are consistent with the conclusion that about a fifth of the sample had constructed the CP but not the cardinal-count concept. In brief, it is unclear whether *meaningful CP-based* success on the how-many task evolves before, or simultaneously with, success on the give-*n* task, which Fuson (1988) presumed required a distinct and more advanced cardinal-count concept.

Issue 2: when do constancy concepts (PL2) develop in relation to the CP and cardinal-count concept (PL1 and PL3, respectively) and each other? The evidence about the issue is sparse and inconclusive. Sarnecka and Gelman (2004) used a transform-set task to investigate children’s understanding of number–word identity. Five or six objects were placed in a box and verbally labeled by the experimenter (e.g., “I’m putting six buttons in this box”). The box was then subjected to a numerically irrelevant action (e.g., shaking or rotation) or a numerically relevant one (adding or subtracting an item). Children were then asked, for instance, “Now how many buttons—is it five or six?” Even two-knowers were successful on the task. Sarnecka and Gelman concluded that a number–word identity-like understanding (somewhat similar to PL2A) develops before meaningful counting (PL1), as assessed by the give-*n* task. In contrast, Condry and Spelke (2008) found that children who did not understand that counting could be used to determine the cardinal value of a collection performed at a chance level on a similar identity-like task.

Four reasons might account for why Sarnecka and Gelman’s (2004) results contradict the logical expectation that children would need to understand the CP (PL1) to conserve the cardinal identity of a collection (PL2A) and the results of Condry and Spelke (2008):

1. They measured the CP indirectly with the give-*n* task, which may underestimate PL1 knowledge.
2. Their transform-set task did not involve children in counting a collection and maintaining the cardinal identity of *their* count—competencies that might reasonably develop after (rather than before) they construct the CP.
3. Some participants’ success may have been due to chance alone. The probability of one or two correct of two large-set identity trials on the transform-set task by guessing alone is 0.50 and 0.25, respectively.
4. Spelke and Condry’s transform-set task, which involved two collections, is arguably more complex and challenging than Sarnecka and Gelman’s task.

Using a compare-set task to measure cardinal equivalence, Sarnecka and Wright (2013) found that only children who were successful on the give-*n* task with larger collections appear to understand what it means for two numbers to be exactly equal (e.g., view a second set as “five” when the set was 1-to-1 correspondence with another set the tester counted and labeled as “five”). They concluded that the CP-knower level (as measured by

the give- n task) was a prerequisite for counting-based cardinal equivalence with larger collections (PL2B).

However, if give- n task actually measures a more advanced (PL3) cardinal-count concept and not CP (PL1), then such an understanding of equivalence is a relatively late development, (i.e., follows PL3). Furthermore, Sarnecka and Wright's (2013) compare-set task may have underestimated cardinal equivalence (PL2B). As counting was disallowed, their task did not directly assess PL2B. Instead, children were shown two pictured, linear collections that were either equal in number in length and number or not, asked whether the two collections were equivalent (control question), told the numerical value of one collection (n), and asked the test question: if the other collection had n or $n + 1$ ($n - 1$). In effect, the task demands of compare-set may have been too challenging for many children of about 3 years of age (Sarnecka & Gelman, 2004). Indeed, about one-third of the trials were discarded because a child did not answer the control question correctly. The test questions required attentiveness and holding the original cardinal value in working memory. The inequivalence trials, in particular, required a relatively advanced verbal-counting skill: knowledge of number-after relations. For example, with five (labeled "five" by the tester) and six, a participant was asked: "Does the (unlabeled) collection have five or six?" This might be confusing to a child who does not yet know that "six" comes after and is (one) more than "five." In brief, it is unclear whether a more straightforward measure of counting-based equivalence and separate measures of the CP and cardinal-count concepts would put this competence between the two latter concepts (i.e., at PL2B as shown in Table 1), after both (i.e., after PL3), or—if the count-cardinal and cardinal-count concepts are not distinct—after the CP (i.e., as Sarnecka & Wright concluded).

1.3 Hypotheses addressed by the present research

Hypothesis 1 (H1) bearing on Issue 1: CP develops before a distinct cardinal-count concept One aim of the present research was to circumvent the methodological limitations of previous efforts to test Fuson's (1988) hypothesis regarding the possible developmental relation between the CP (count-cardinal concept) and the cardinal-count concept. Specifically, the present study served to test H1:

An understanding of the CP, as assessed by a battery of relatively meaningful how-many tasks and a CP-application task, develops before its inverse, the cardinal-count concept, as gauged by the give- n task.

The how-many task used in the present study is a game-based version of Schaeffer et al.'s (1974) hiding task and specifically designed to address the concerns that a how-many task can underestimate competence (Gelman, 1993; Le Corre et al., 2006). In addition to the how-many task with linear arrays, a more difficult how-many task with haphazard arrays was also administered. Unlike the linear version, the non-linear version entailed more than minimal effort to keep track of which items have been counted and which need to be counted (Beckwith & Restle, 1966; Potter & Levy, 1968; Schaeffer et al., 1974). Most importantly, a "CP-application task" was administered to provide an indicator of whether a child was merely using a last-word rule by rote or the CP. This cardinal-identity task, unlike Sarnecka and Gelman's (2004) transform-set task, required a child to determine the

cardinal value of a collection by counting the collection themselves and then applying this information in a meaningful manner.

H2 bearing on a methodological implication of Issue 1: give-*n* performance underestimates CP understanding The give-*n* task is widely used in developmental psychology to assess the CP. However, if Fuson's (1988) hypothesis is correct and a distinct and developmentally more advanced cardinal concept, namely, the cardinal-count concept, is required for successfully counting out a specified number of items (i.e., success on the give-*n* task), then this popular operational definition of the CP may underestimate CP competence. The second research aim, then, was to evaluate the validity of the give-*n* task as a measure of the CP. This involved testing H2:

Performance on the give-*n* task will be significantly lower than that on a battery of how-many tasks and a CP-application task assessing a meaningful understanding of the CP.

H3 bearing on Issue 2: CP→constancy concepts→cardinal-count concept The third aim of the present research was to address Issue 2—how cardinal identity and equivalence were related to other cardinality concepts and each other. In contrast to Sarnecka and Gelman's (2004) conclusion that cardinal identity develops before the CP as gauged by the give-*n* task (but cf. Condry & Spelke, 2008), successful application of this principle with self-counted collections should emerge after the CP or *meaningful* counting (PL1), which Sarnecka and Gelman did not assess. In contrast to Sarnecka and Wright's (2013) finding, cardinal equivalence—as assessed by a meaningful CP measure that involves self-counted collections and fewer task demands—may develop before CP-knower level as gauged by the give-*n* task (which may actually assess the cardinal-count concept). The present research tested the hypothesized developmental order of cardinality concepts indicated in Table 1 (H3):

PL1 (CP) → PL2A (cardinal identity) → PL2B (cardinal equivalence) → PL3 (cardinal-count concept).

2 Method

2.1 Participants

A total of 23 children ($M = 49.4$ months, $SD = 3.95$ months, range = 41.5–54.9 months) participated, including nine 3-year-olds (five girls) and fourteen 4-year-olds (six girls). Children were recruited from preschool programs serving predominantly middle-class communities in Taiwan. Informed consent was obtained for 24 children, but one 4-year-old girl did not complete the second of two sessions because of a serious illness.

2.2 Tasks and materials

Task 1: how many—linear With the *Hidden Stars* game, children were presented a card with linear quantity of 5 and then 6 stars one-inch stars ½-inch apart and asked to count the stars. Then to create a purpose for applying the CP, the tester hid the stars and asked the child how many stars were hidden. Correct 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 (was either accurate or involved only a single minor slip). Incorrect responses included answering the how-many question with a different number, simply repeating the count, or not responding (range=0 to 2 correct). A practice trial involving 3 linear stars served to introduce the task demands.

Task 2: give- n *Feeding Ravenous* entailed providing a pile of 10 chips (“cookies”) and asking a child to put 5 and then 6 cookies in the Muppet’s mouth. Correct was defined as *counting out* the requested amount correctly or making a single minor error but honoring the cardinal-count concept (i.e., labeling the last item produced with the requested number). Responses otherwise were scored as incorrect (range=0 to 2). For example, for the give-6 trial, accidentally producing five or seven items while counting “1, 2, 3, 4, 5, 6” was scored as correct. A child who simply grabbed some items, whether six or not, was also scored as 0, because the cardinal-count concept was not applied. A practice trial involving 3 served to familiarize participants with the task.

Task 3: how many—haphazard The procedure and scoring of the *Hidden Chips* game were like the *Hidden Stars* game but involved a practice trial of 4 and tests trials of 6 and then 5. The trials were similar to the give- n task in two ways: the arrays were haphazardly arranged and consisted of moveable objects (chips).

Task 4: cardinal identity In the *Hidden Cookies* game, a child first generated the cardinal number for a collection of 2 and then 3 (practice trials) and 6 and then 5 (test trials) by counting. The collection was then immediately hidden to prevent re-counting, and an action was performed. The child was then asked whether the action affected the total (addition or subtraction of 1) or not (irrelevant physical action). The first practice trial (tester tapped a covered collection of 3), trial 1 (tester waved a hand over a covered collection of 5 chips), and trial 4 (tester puffed on a covered collection of six cookies) involved conservation of cardinal identity. The second practice trial (tester added a chip to a collection of two), trial 2 (tester removed an item from a collection of six chips), and trial 3 (tester added a chip to a collection of five) involved an arithmetic transformation and served as control trials (i.e., to detect a response bias of always stating the last number word in a count). After an operation, a child was asked, “How many cookies am I hiding?”.

On *cardinal-identity trials*, correct (scored as 1 point) was defined as a response to the how-many question that matched the last number word used in the count whether correct or not. (In fact, all participants who were scored as correct on a cardinal-identity trial counted accurately.) On *transformation or control trials*, a response to the how-many question other than the last number word of the original (pre-transformation) count was considered an appropriate response (i.e., indicated the lack of a response bias); 0 points were deducted for an appropriate response. Responding with the last number word of the original count was considered inappropriate (i.e., an indicated response bias) and scored as -1 (allowable range=0 to 2 points).

Task 5: cardinal equivalence The *Number Clues* game required a child to use counting to determine which of three collections were equal in number to a counted collection of dissimilar shapes and arrangements (i.e., recognize that two collections with the same cardinal value are numerically equivalent despite differences in physical appearance). The game entailed presenting three boxes, one of which hid “gold.” Each box had a code of 3 to

7 dots presented in a regular (dominoes) array. A clue in the form of a linear array of 3 (practice trial), 6 (trial 1), or 5 (trial 2) squares was presented along with the instructions: “Here is your clue. Count these squares.” After the child counted the collection, the tester said: “Which box do you think is hiding the gold?” Choosing a collection that matched the cardinal value of the clue count was scored as correct whether the latter was correct or not (range = 0 to 2 correct).

2.3 Procedure

Tasks were administered over two sessions in the following prescribed order: tasks 1 and 2 in session 1 and tasks 3, 4, and 5 in session 2. Session 1 took 3 or 4 minutes; session 2 took 5 or 6 minutes. Session 2 was administered about 1 week after session 1. The correct procedure and answer were modeled for children who missed the practice trial.

2.4 Analyses

H1 A child’s performance on the two how-many tasks (tasks 1 and 3) and the cardinal-identity task (task 4) was used to create a composite CP score to identify accurate and meaningful CP knowledge, accurate application of a last-word rule, or no CP knowledge. A child’s score on the give- n task (task 2) served to indicate whether a child understood the cardinal-count concept. A Yule’s Q , a nonparametric test of the association between two dichotomous ordinal variables, served as an index of effect size and ranges from -1 to $+1$, with 0 indicating no effect (Bakeman & Quera, 2011). An index of $+1$ indicates a perfect association and the ideal support for H1—the “CP-priority hypothesis” that an understanding of the CP develops before an understanding of the cardinal-count concept. A perfect association occurs if one variable is always equal to or higher than another—that is, if all the data are distributed among three cells in a 2×2 table: successful on variable A and either not successful or successful on variable B and not successful on both variables A and B (Dixon & Moore, 2000). Note that the remaining cell (unsuccessful on variable A but successful on variable B) should be 0. A Yule’s $Q > 0.70$ indicates a very strong association; 0.50 to 0.69, a substantial association; 0.30 to 0.49, a moderate association; and 0 to 0.29, no or a negligible association. Developmental synchrony—the simultaneous development of the two concepts—would ideally be supported if the data were distributed between two cells in a 2×2 matrix: successful on both concepts and unsuccessful on both. Sarnecka and Carey’s (2008) alternative hypothesis that the concepts are not distinct would predict results indistinguishable from a developmental synchrony hypothesis.

H2 To test whether children’s give- n score was equivalent to their composite CP score (H2), a test for equality of partially overlapping frequencies (McNemar’s test) was used to compare whether success on the former but non-success on the latter was equal to the reverse. This test, not a chi-square, is appropriate for two non-independent, dichotomous traits (i.e., a 2×2 contingency table involving repeated measures of the same sample; Darlington, 1974).

H3 As the assumptions of parametric test were not met, the analysis for addressing H3 involved a nonparametric test. In addition to significance level, effect size is reported as a Cohen’s d . Unlike significance level, effect size bears on practical significance and indicates statistical magnitude of the effect (Lipsey et al., 2012; Wilkinson et al., 1999). The

following guidelines from Cohen (1988, 1992) apply: $d=0.2$, 0.5, and 0.8 indicate a small, medium, and large effect size, respectively.

3 Results

3.1 H1: CP develops before a distinct cardinal-count concept

Table 2 summarizes the participants' CP composite score. Nine participants were completely unsuccessful tasks 1, 3, and 4 (cell A in Table 2)—that is, did not apply even the last-word rule. Among the three children who were partially successful on the how-many tasks (tasks 1 and 3), two were unsuccessful in applying their CP knowledge on the cardinal-identity task (task 4; cell D in Table 2), indicating they may have been using only the last-word rule on the how-many task. One child went from being completely unsuccessful on the initial how-many task to completely successful on the second how-many task and was subsequently successful on the cardinal-identity task (cell F). These results are consistent with the child constructing the CP during the study. Indicative of CP knowledge, the remaining 11 participants were consistently successful on the how-many tasks and had at least some success on the cardinal-identity task (cells H and I in Table 2). In sum, 12 of 14 children who were successful on the how-many tasks appeared to understand the CP.

Table 3 summarizes the relation between participants' CP composite score and their performance on the give- n task. Note that consistent with a CP-priority hypothesis, *all* the data fell in cells in which CP development is either equal to or higher than that of the cardinal-count concept (i.e., cells A, D, G, H, J, K, M, N, and O)—if the former developed earlier and served as a necessary condition for the latter. Note that all six developmentally unambiguous cases in Table 3 (cells G, J, and M) are consistent with H1—a CP-priority (Fuson's 1988) hypothesis—and inconsistent with a reverse-priority hypothesis or

Table 2 Relative success on the two how-many tasks and the cardinal-identity task ($n=23$)

How-many tasks (Tasks 1 & 3)	Cardinal-identity task (Task 4, meaningful application of the CP)		
	Unsuccessful (0 correct)	Partially successful (1 correct)	Successful (2 correct)
Successful (3 or 4 correct)	0 ^G	2 ^H	9 ^I
Partially successful (2 correct)	2 ^D	0 ^E	1 ^F
Unsuccessful (0 or 1 correct)	9 ^A	0 ^B	0 ^C

The scores on both tasks were used to create a composite CP score:

- Strong evidence of meaningful CP knowledge = cell I
- Evidence of emerging secure CP knowledge = cell H
- Evidence of learning the CP = cell F
- Evidence of last-word rule = cell D
- Evidence not consistent with knowing the last-word rule or the CP = cell A

Table 3 CP knowledge (the composite score of the two how-many and the cardinal-identity tasks) × give-*n* task success

		Give- <i>n</i> task		
		0 correct	1 correct	2 correct
Exhibited CP success/knowledge	Strong evidence of secure CP knowledge—consistently successfully on the how-many and cardinal-identity tasks (cell I in Table 2)	4 ^M	0 ^N	5 ^O
	Evidence of emerging secure CP knowledge—consistently successful on how-many tasks and partially successful on cardinal-identity task (cell H in Table 2)	1 ^J	1 ^K	0 ^L
	Appeared to learn the CP—unsuccessful and then successful on how-many tasks and partially successful on cardinal-identity task (cell F in Table 2)	1 ^G	0 ^H	0 ^I
Did NOT exhibit CP success/knowledge	Inconsistently used or learned the last-word rule—partially successful on how-many tasks and unsuccessful on cardinal-identity task (cell D in Table 2)	2 ^B	0 ^E	0 ^F
	Little or no evidence of CP knowledge—unsuccessful on how-many and cardinal-identity tasks (cell A in Table 2)	9 ^A	0 ^B	0 ^C

A conservative criterion for give-*n* (GN) success was defined as correct on both trials (2 correct). Yule's *Q* (success on both tasks = cells I + L + O = 5; success on CP but not GN = cells G + H + J + K + M + N = 7; success on GN but not CP = cells C + F = 0; and unsuccessful on both = cells A + B + D + E = 11) = + 1.0, *p* = .003. A liberal criterion for GN success was defined as correct on at least one trial (1 or 2 correct). Yule's *Q* (success on both tasks = cells H + I + K + L + N + O = 6; success on CP but not GN = cells G + J + M = 6; success on GN but not CP = cells B + C + E + F = 0; and unsuccessful on both = cells A + D = 11) = + 1.0, *p* < .001

a synchronous-development (Sarnecka & Carey's 2008) hypothesis.¹ Specifically, the two children who were in the process of constructing the CP and the four who exhibited secure knowledge of the CP had no success on the give-*n* task. As the note in Table 3 indicates, there was a perfect and statistically significant positive correlation between the level of CP knowledge and cardinal-count understanding, whether a conservative criterion (correct on both give-*n* trials) or a liberal criterion (correct on at least one give-*n* trial) was used.

Moreover, the evidence also appears inconsistent with a developmental synchrony hypothesis or Sarnecka & Carey's, 2008 non-distinct hypothesis. Among all cells in Table 3 indicating a developmental transition (other than cells A and D, which neither possible construct has begun to develop, or cell O, which represents the completed development of the two possible constructs), six cases are *inconsistent* with the simultaneous development of two distinct concepts or a non-distinct concept. The seventh case (cell K) is consistent with Sarnecka and Carey's (2008) non-distinct hypothesis, but it is also consistent with prior development of the CP, prior development of cardinal-count concept, or the simultaneous development of two distinct concepts.

3.2 H2: give-*n* performance underestimates CP understanding

In Table 3, using a liberal criterion for give-*n* success, success on the CP knowledge battery but non-success on the give-*n* task (cells M, J, and G) equals 6, success on both (cells N, O, K, and L) is 6; non-success on both (cells D and A) is 11, and non-success on the CP knowledge battery but success on the give-*n* task equals 0. Consistent with H2, the McNemar's test indicates that the CP knowledge score was significantly better than that of the give-*n* ($p=0.0156$, one-tailed).

3.3 H3:CP→constancy concepts→cardinal-count count

The results were generally consistent with H3—in accord with the HLP summarized in Table 1. Participants performed best on the how-many-haphazard and how-many-linear tasks ($M=1.261$, $SD=0.8643$, and $M=1.043$, $SD=0.9760$, respectively), most poorly on the give-*n* task ($M=0.522$, $SD=0.9458$), and somewhere in between for the two number constancy tasks ($M=0.957$, $SD=0.8299$ for cardinal identity, and $M=0.870$, $SD=0.626$ for cardinal equivalence). A Friedman test indicated that success on the five cardinality tasks differed significantly ($\chi^2=16.128$ [4], $p=0.003$). As expected, performance on the two how-many did not differ significantly ($Z=-1.508$, two-tailed $p=0.132$). Performance on the how-many-linear and how-many-haphazard tasks were each significantly and substantially (as measured by a medium or large effect size) better than that on the give-*n* task ($Z=-2.585$, two-tailed $p=0.01$, $d=0.57$, and $Z=-3.153$, two-tailed $p=0.002$, $d=0.86$, respectively). In three of four cases, performance on the how-many-linear and how-many-haphazard tasks was each significantly and substantially (as measured by a small or medium effect size) better than that on the cardinal-identity task ($Z=-1.269$, two-tailed $p=0.204$, $d=0.09$, and $Z=-2.021$, two-tailed $p=0.043$, $d=0.36$, respectively) and the

¹ The cases in cells A, D, E, H, K, and O are developmentally ambiguous, because it is consistent with a CP-priority (Fuson's 1988 hypothesis), reverse-priority, or synchrony (Sarnecka and Carey's 2008 non-distinct) hypothesis.

Table 4 Relative success on CP understanding and reliable cardinal equivalence knowledge ($n=23$)

CP understanding (CP composite score)	Understanding of cardinal equivalence	
	No (0 or 1 correct on Task 5)	Yes (2 correct on Task 5)
Yes (Cells F, H, & I in Table 2)	8 ^C	4 ^D
No (Cells A & D in Table 2)	10 ^A	1 ^B

cardinal-equivalence task ($Z = -3.116$, two-tailed $p = 0.002$, $d = 0.21$, and $Z = -3.477$, two-tailed $p = 0.001$, $d = 0.52$, respectively).

Moreover, as Table 4 illustrates, an understanding of the CP as measured by a composite CP score clearly emerged before reliable cardinal equivalence, as measured by a substantial association (effect size). Specifically, with four successful scores on both tasks (cell D of Table 4), eight scores indicating priority of CP knowledge (cell C), only one score indicating priority of cardinal equivalence (cell B), and 10 unsuccessful on both concepts (cell A) Yule's Q was 0.667 and marginally significant (one-tailed $p = 0.098$).

The two number constancy tasks also did not differ significantly in difficulty from each other ($Z = -1.602$, two-tailed $p = 0.109$). However, if reliable performance on each task is considered (a score of 2 = reliably successful versus a score of 0 or 1 = unsuccessful), then 12 participants were unsuccessful on both; only 1 was unsuccessful on cardinal identity but reliably successful on cardinal equivalence; 6 were reliably successful on cardinal identity but unsuccessful on cardinal equivalence, and 4 were reliably successful on both tasks ($Q = 0.778$, approximate $p = 0.061$). Performance on the cardinal-identity task and cardinal-equivalence task was significantly and substantially better than that on the give- n task ($Z = -2.845$, two-tailed $p = 0.004$, $d = 0.51$, and $Z = -2.000$, two-tailed $p = 0.046$, $d = 0.47$, respectively). Competence began to emerge (1 point) or was achieved (2 points) on the cardinal-identity consistently before such development on the give- n task: 11 participants were unsuccessful on both tasks; 5 had at least some success on cardinal-identity task but not on give- n task; none exhibited the reverse pattern of development; and 7 had at least some success on both tasks ($Q = 1.000$, $p < 0.001$). The same was true for cardinal equivalence: 5 participants were incorrect on both; 11 had at least some success on the cardinal-equivalence task but not on give- n task; none exhibited the reverse pattern of development; and 7 had at least some success on both tasks ($Q = 1.000$, $p = 0.015$).

An unanticipated but interesting finding emerged from the data. Five children were exactly correct on the two control trials of the cardinal-identity task involving (mentally) taking one from six and adding one to six. All five children appeared to understand both the CP, as indicated by CP composite score, and the cardinal-count concept, as indicated by reliable success on the give- n task.

4 Discussion

4.1 H1: CP develops before a distinct cardinal-count concept

The results are consistent with Fuson's (1988) hypothesis that an *understanding* of the CP (PL1) emerges before (rather than simultaneously with) that of the cardinal-count concept

(PL3) and that the former is the developmental prerequisite (necessary condition) of the latter. Specifically, the statistically significant and perfect Yule's Q result indicate that meaningful CP knowledge was always equal to or higher than (i.e., developed before) success on the give- n task, which requires the cardinal-count concept. The present results are also consistent with Mou et al.'s (2021) conclusion that distinct concepts underlie how-many and give- n performance.

It has long been recognized that success on the how-many tasks precedes success on the give- n task. Two widely recognized reasons for this disparity have been (a) children can use the last-word rule learned by rote to create the appearance of success on the former but not the latter and (b) greater non-conceptual demands of the give- n task. It is unlikely the first reason accounts for the present results, because success on the how-many task had to be coupled with success on a task that required a child to apply the CP. Although forgetting the requested number or failing to match a count to the requested number because of a working memory overload cannot be discounted, qualitative analyses of the five cases of prior development of the CP (cells J and M in Table 3) suggest a plausible alternative hypothesis: their lack of success on the give- n was due to not understanding the rationale for the counting-out procedure.²

- The cell J child, who appeared to be in the process of consolidating the CP, and two cell M children, who exhibited a secure CP knowledge, made no effort to count out items but simply grabbed some or all 10 items available on both give- n trials. Like five of the cell A children who were unsuccessful on both tasks, these three CP-knowers may have relied on a grabbing strategy, because they had not constructed the cardinal-count concept and did not understand how counting could be used in the service of producing a requested set.
- A cell M child, who exhibited strong evidence of understanding the CP, made the classic no-stop error of counting all the available items for both give- n trials, without any effort to correct herself. Like one cell A child who made such an error on both give- n trials and one cell D child (a last-word rule user) who did so on one trial, this CP-knower may not have stopped the counting-out process, because she did not understand the cardinal-count concept.
- For the give-5 trial, a cell M child counted past “five” to produce six and made no effort to correct herself. On the give-6 trial, she counted out four items. In each case, the girl violated the cardinal-count concept by not stopping the counting-out process at the number requested. This CP-knower (like a cell A child and a cell D child who each twice counted past the requested number) may have made such an error because she did not know the cardinal-count concept.

It could be argued that because the tasks were administered in a fixed order, order effects might confound the results. Although some children's familiarity and comfort level might increase with testing, this might be counterbalanced by increased fatigue by other participants. Nevertheless, future research should control for possible order effects when testing a hypothesis like H1 (and H2 and H3 as well).

² The child who appeared to construct an understanding of the count-cardinal concept but had no success on the give- n task (cell G in Table 3) is not included in this analysis. Evidence of the former was collected in session 2; evidence of the latter was in session 1. So, it is unclear whether the girl would also have constructed the cardinal-count concept and been successful on the give- n task in session 2.

Like Frye et al.'s (1989) and Sarnecka and Carey's (2008) data, the present results indicated that young children's performance on the how-many tasks and give-*n* are not equivalent for larger collections of 5 and 6. Unlike their inference that this performance gap was largely due to a meaningless last-word rule, the present analysis indicated that 86% of those who accurately responded to how-many questions also appeared to apply the CP meaningfully on the cardinal-identity task.

4.2 H2: give-*n* performance underestimates CP understanding

When use of the last-word rule as basis of success on the how-many task is eliminated, which prevents over-estimating CP competence, performance on the how-many task was still significantly and substantially better than that on the give-*n* task. This result indicates the give-*n* task underestimates understanding of the CP.

4.3 H3: CP→constancy concepts→cardinal-count concept

The present results accord with the proposed developmental order of concepts in Table 1. Consistent with Condry and Spelke's (2008) results, children performed significantly and substantively better on the how-many-haphazard task, indicative of knowing the last-word rule or perhaps even the count-cardinal concept (PL1), than on the cardinal-identity task with collections that children counted for themselves (PL2A). However, the data summarized in Table 2 are not sufficient to contradict Sarnecka and Gelman's (2004) finding that cardinal identity develops before the CP, and further research is needed to clarify the developmental order of these concepts. Reliable success with cardinal identity outpaced that with cardinal equivalence (PL2B) at a marginally significant but (as indicated by effect size) strongly substantive level. Cardinal equivalence developed after a meaningful understanding of the CP but not before it, as Sarnecka and Wright (2013) found when the concept was measured by the give-*n* task (PL3).

It could be argued that cardinal-equivalence task used in the present study overestimates knowledge because the probability of guessing the correct answer is one-in-three. This would have worked against finding cardinal equivalence more challenging than understanding the CP or cardinal identity but may have contributed to why cardinal equivalence performance appeared better than for the give-*n* task. Further research with a task that reduces the possibility of false positives is needed to examine whether cardinal equivalence actually develops before the cardinal-count concept.

On the cardinal-identity task, five children appeared to understand that adding one to a collection means moving forward exactly one word in the counting sequence and removing one means moving backward exactly one word (i.e., exhibited implicit knowledge of the successor function). All five were also completely successful on the how-many and the cardinal-identity tasks—that is, exhibited an understanding of the CP (cf. Sarnecka & Carey, 2008). These children were also completely successful on the give-*n* trials, which Fuson (1988) hypothesizes requires an understanding of the more advanced cardinal-count concept. These results, then, are consistent with the growing evidence that the successor function develops relatively late, after the CP (see Schneider et al., 2021, for a review). They are inconsistent with Carey's (2004) bootstrapping hypothesis that children use analogical reasoning between their cardinal understanding of small numbers and the structure of the count list to acquire the CP—that becoming a CP-knower entails learning the successor function. Instead, children may induce the CP from counting small collections they

can subitize and recognizing that the outcome of the one-to-one counting process consistently matches the number they “see” (Paliwal & Baroody, 2020).

5 Conclusions

In terms of theoretical implications, although the study clearly has limited external validity and needs to be replicated with a wider range of materials, more trials (including those just beyond five), and a larger and more chronologically, demographically, and linguistically diverse sample, its preliminary results support the HLP outline in Table 1. For example, future research should consider the caution that children who can successfully create a set of 5 may not be able to do so with larger sets—as was the case for two participants in the present study (Posid & Cordes, 2018; Sella et al., 2021). Such children may not know a cardinal-count concept but “successfully” create a set of five any way using a subitizing-based putting-out strategy instead. Specifically, if they can subitize 4, children may put out items one at a time until they see “four” and then put out another item because they know “five” is more than “four.” To minimize failures due to performance factors, future research could also use a more direct measure of the cardinal-count concept than the *give-*n** task. For instance, a *stop-at-*n** task entails instructing a child to stop a counting-out process of another at a specified number. The tester counts out items—relieving a child of the procedural demands of the counting-out process, allowing the child to focus on applying the cardinal-count concept.

The present results are consistent with a follow-up experiment (Baroody et al., 2021). Children who had not achieved PL1 (the CP) or PL3 (the counting-out procedure and its rationale the cardinal-count concept) were randomly assigned to two interventions. One was based on the HLP in Table 1; a comparison intervention focused on other aspects of cardinality but involved the same PL3 training. The HLP-based intervention resulted in significantly and (as measured by effect size) substantially greater learning on the cardinal-count concept (as measured by the *stop-at-*n** task) and procedural fluency (as measured by *give-*n** task). As the present results and those of the follow-up experiment indicate that PL1 and PL3 may involve distinct competencies (e.g., different conceptual bases, types of mapping, and demands on working memory) and the former emerges before the latter for collections beyond the subitizing range, it does not make sense to refer to both competencies with the same terminology—CP-knower level. The CP-knower level should refer only to PL1.

The preliminary findings regarding the developmental order of the constancy concepts—cardinal identity (PL2A in Table 1) and cardinal equivalency (PL2b)—need to be replicated. With additional evidence, it might well make sense to list each of these extensions of the CP as separate levels as shown in Table 1.

In terms of methodological implications, the present results are consistent with the growing concern that the *give-*n** task may underestimate (pre-counting and) counting-based cardinality knowledge (Barner & Bachrach, 2010; Krajcsi, 2021; Mou et al., 2021; Sella et al., 2021; Wagner et al., 2019). If future results confirm that the *give-*n** task requires the more advanced cardinal-count concept, assessing the CP-knower level might better involve a combination of tasks such as the *how-many* task and a *how-many* application task (e.g., the cardinal-identity task) and using a composite cardinality score as in the present study. Such a procedure provides a compromise between overestimating CP knowledge (as when

the how-many task is used alone) and overestimating this knowledge (as when the give- n task is used alone).

In terms of educational implications, if present results supporting the HLP in Table 1 can be replicated, then it can serve as the basis for sequencing early mathematics counting and number instruction. School reformers and educational researchers have embraced instruction based on HLPs as a potentially important tool for reform (Frye et al., 2013; Lobato & Walters, 2017; Maloney et al., 2014; Sarama & Clements, 2009; Shavelson & Karplus, 2012; Simon, 1995). Baroody et al. (2021) found that most participants who first learned the CP benefitted from instruction on the cardinal-count concept and the counting-out procedure, whereas all those who did not previously construct an understanding of the CP did not.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10649-022-10153-5>.

Author contribution Dr. Arthur J. Baroody was the primary contributor to the study conception and design. Material preparation, data collection, and analysis were performed by Dr. Menglung Lai. The first author wrote the first draft of the manuscript, and the second author commented on previous versions of the manuscript. Both authors read and approved the final manuscript.

Funding The preparation of this report was supported by the Institute of Education Science [grant number R305A150243] and the National Science Foundation [grant number 1621470] 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.

Availability of data and material The data and materials are available upon request to an author.

Code availability Not applicable.

Declarations

Ethics approval In Taiwan, researchers need only to obtain parent consent.

Consent to participate Informed parental consent was obtained before testing.

Consent for publication Not applicable.

Conflict of interest Not applicable.

References

- Barner, D., & Bachrach, A. (2010). Inference and exact numerical representation in early language development. *Cognitive Psychology*, 60, 40–62. <https://doi.org/10.1016/j.cogpsych.2009.06.002>
- Baroody, A. J., Clements, D. H., & Sarama, J. (2021). Lessons learned from 10 experiments that tested the efficacy and assumptions of hypothetical learning trajectories *Education Science*, 12. <https://doi.org/10.3390/educsci12030195>
- 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. Saracho (Eds.), *Handbook of research on the education of young children* (pp. 187–221). Erlbaum.
- Baroody, A. J., & Purpura, D. J. (2017). Early number and operations: Whole numbers. In J. Cai (Ed.), *Compendium for research in mathematics education* (pp. 308–354). National Council of Teachers of Mathematics.

- Bakeman, R., & Quera, V. (2011). Sequential analysis and observational methods for the behavioral sciences. *Cambridge University Press*. <https://doi.org/10.1017/CBO9781139017343>
- Beckwith, M., & Restle, F. (1966). Process of enumeration. *Psychological Review*, 73(5), 437–444. <https://doi.org/10.1037/h0023650>
- 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>
- Carey, S. (2004). Bootstrapping & the origin of concepts. *Daedalus*, 133, 59–68. <https://doi.org/10.1162/001152604772746701>
- Carey, S., & Barner, D. (2019). Ontogenetic origins of human integer representations. *Trends in Cognitive Sciences*, 23(10), 823–835. <https://doi.org/10.1016/j.tics.2019.07.004>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Routledge. ISBN 1–134–74270–3.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159.
- Condry, K. F., & Spelke, E. S. (2008). The development of language and abstract concepts: The case of natural number. *Journal of Experimental Psychology: General*, 137, 22–38. <https://doi.org/10.1037/0096-3445.137.1.22>
- Darlington, R. B. (1974). *Radicals and squares and other statistical procedures for the behavioral sciences*. Logan Hill Press.
- Dixon, J. A., & Moore, C. F. (2000). The logic of interpreting evidence of developmental ordering: Strong inference and categorical measures. *Developmental Psychology*, 36, 826–834. <https://doi.org/10.1037/0012-1649.36.6.826>
- Fischer, J. P. (1992). Subitizing: The discontinuity after three. In J. Bideaud, C. Meljac, & J. P. Fischer (Eds.), *Pathways to number* (pp. 191–208). Erlbaum.
- Frye, D., Baroody, A. J., Burchinal, M., Carver, S. M., Jordan, N. C., & McDowell, J. (2013). *Teaching math to young children: A practice guide*. Washington, DC: National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education. <http://ies.ed.gov/ncee/wwc/practiceguide.aspx?sid=18>.
- Frye, D., Braisby, N., Lowe, J., Maroudas, C., & Nicholls, J. (1989). Children's understanding of counting and cardinality. *Child Development*, 60, 1158–1171. <https://doi.org/10.2307/1130790>
- Fuson, K. C. (1988). *Children's counting and concepts of number*. Springer-Verlag.
- Gallistel, C. R., & Gelman, R. (2000). Non-verbal numerical cognition: From reals to integers. *Trends in Cognitive Sciences*, 4(2), 59–65. [https://doi.org/10.1016/s1364-6613\(99\)01424-2](https://doi.org/10.1016/s1364-6613(99)01424-2)
- Gelman, R. (1993). A rational-constructivist account of early learning about numbers and objects. Advances in research theory. In D. L. Medin (Ed.), *The Psychology of Learning and Motivation* (pp. 61–96). Academic Press.
- Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkman, J. (1949). The discrimination of visual number. *American Journal of Psychology*, 62, 498–525. <https://doi.org/10.2307/1418556>
- Klahr, D., & Wallace, J. G. (1976). *Cognitive development: An information-processing view*. Erlbaum.
- Krajcsi, A. (2021). Follow-up questions influence the measured number knowledge in the give-a-number task. *Cognitive Development*, 57, 100968. <https://doi.org/10.1016/j.cogdev.2020.100968>
- Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition*, 105, 395–438. <https://doi.org/10.1016/j.cognition.2006.10.005>
- Le Corre, M., & Carey, S. (2008). Why the verbal counting principles are constructed out of representations of small sets of individuals: A reply to Gallistel. *Cognition*, 107, 650–662. <https://doi.org/10.1016/j.cognition.2007.09.008>
- 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://doi.org/10.1016/j.cogpsych.2005.07.002>
- Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., Roberts, M., Anthony, K. S., & Busick, M. D. (2012). *Translating the statistical representation of the effects of education interventions into more readily interpretable forms*. IES National Center for Special Education Research, Institute of Education Sciences.
- 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>
- Nieder, A. (2017). Number faculty is rooted in our biological heritage. *Trends in Cognitive Sciences*, 21(6), 403–404.
- Paliwal, V., & Baroody, A. J. (2020). Cardinality principle understanding: The role of focusing on the subitizing ability. *ZDM-Mathematics Education*, 52(4), 649–661. <https://doi.org/10.1007/s11858-020-01150-0>
- Potter, M. C., & Levy, E. I. (1968). Spatial enumeration without counting. *Child Development*, 39(1), 265–272. <https://doi.org/10.2307/1127377>
- Posid, T., & Cordes, S. (2018). How high can you count? Probing the limits of children's counting. *Developmental Psychology*, 54(5), 875–889. <https://doi.org/10.1037/dev0000469>
- Resnick, L. B., & Ford, W. W. (1981). *The psychology of mathematics for instruction*. Erlbaum.
- 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>
- Sarnecka, B. W., & Gelman, S. A. (2004). Six does not just mean a lot: Preschoolers see number words as specific. *Cognition*, 92, 329–352. <https://doi.org/10.1016/j.cognition.2003.10.001>
- Sarnecka, B. W., & Wright, C. E. (2013). The idea of an exact number: Children's understanding of cardinality and equinumerosity. *Cognitive Science*, 37, 1493–1506. <https://doi.org/10.1111/cogs.12043>
- Schaeffer, B., Eggleston, V. H., & Scott, J. L. (1974). Number development in young children. *Cognitive Psychology*, 6, 357–379. [https://doi.org/10.1016/0010-0285\(74\)90017-6](https://doi.org/10.1016/0010-0285(74)90017-6)
- Schneider, R. M., Sullivan, J., Guo, K., & Barner, D. (2021). What counts? Sources of knowledge in children's acquisition of the successor function. *Child Development*, 92(4), e476–e492. <https://doi.org/10.1111/cdev.13524>
- Sella, F., Slusser, E., Odic, D., & Krajcsi, A. (2021). The emergence of children's natural number concepts: Current theoretical challenges. *Child Development Perspectives*. <https://doi.org/10.1111/cdep.12428>
- Shavelson, R. J., & Karplus, A. (2012). Reflections on learning progressions. In A. C. Alonzo & A. W. Gottwals (Eds.), *Learning progressions in science: Current challenges and future directions* (pp. 13–26). Sense.
- 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>
- Von Glasersfeld, E. (1982). Subitizing: The role of figural patterns in the development of numerical concepts. *Archives de Psychologie*, 50, 191–218.
- Wagner, K., Chu, J., & Barner, D. (2019). Do children's number words begin noisy? *Developmental Science*, 22(1), e12752.
- Wilkinson, L., APA Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54, 594–604. <https://doi.org/10.1037/0003-066X.54.8.594>
- Wynn, K. (1990). Children's understanding of counting. *Cognition*, 36, 155–193. [https://doi.org/10.1016/0010-0277\(90\)90003-3](https://doi.org/10.1016/0010-0277(90)90003-3)
- Wynn, K. (1992). Children's acquisition of the number words and the counting system. *Cognitive Psychology*, 24, 220–251. [https://doi.org/10.1016/0010-0285\(92\)90008-P](https://doi.org/10.1016/0010-0285(92)90008-P)

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.