

The Development and Assessment of Early Cardinal-Number Concepts

Arthur J. Baroody¹ , Kelly S. Mix² , Gamze Kartal³ , Meng-lung Lai⁴ 

[1] Department of Curriculum & Instruction, University of Illinois at Urbana-Champaign, Champaign, IL, USA. [2] Department of Human Development and Quantitative Methodology, University of Maryland, College Park, MD, USA. [3] Department of Educational Psychology, University of Illinois at Urbana-Champaign, Champaign, IL, USA. [4] Department of Early Childhood Education, National Chiayi University, Chiayi, Taiwan.

Journal of Numerical Cognition, 2023, Vol. 9(1), 182–195, <https://doi.org/10.5964/jnc.10035>

Received: 2022-08-08 • Accepted: 2022-12-06 • Published (VoR): 2023-03-31

Handling Editor: Tali Leibovich-Raveh, University of Haifa, Haifa, Israel

Corresponding Author: Arthur J. Baroody, University of Illinois at Urbana-Champaign, College of Education, 311 Education Building, 1310 South Sixth Street, Champaign, IL 61820 USA. E-mail: baroody@illinois.edu

Supplementary Materials: Data [[see Index of Supplementary Materials](#)]



Abstract

Number-recognition tasks, such as the how-many task, involve set-to-word mapping, and number-creation tasks, such as the give-n task, entail word-to-set mapping. The present study involved comparing sixty 3-year-olds' performance on the two tasks with collections of one to three items over three time points about 3 weeks apart. Inconsistent with the sparse evidence indicating equivalent task performance, an omnibus test indicated that success differed significantly by task (and set size but not by time). A follow-up analysis indicated that the hypothesis that success emerges first on the how-many task was, in general, significantly superior to the hypothesis of simultaneous development. It further indicated the how-many-first hypothesis was superior to a give-n-first hypothesis for sets of three. A theoretical implication is that set-to-word mapping appears to develop before word-to-set mapping, especially in the case of three. A methodological implication is that the give-n task may underestimate a key aspect of children's cardinal understanding of small numbers. Another is that the traditional give-n task, which requires checking an initial response by one-to-one counting, confounds pre-counting and counting competencies.

Keywords

assessment, cardinality development, early childhood, give-n, how many, subitizing

Researchers have used a variety of measures to assess children's verbal-based cardinal-number knowledge—understanding that a number word represents a specific number of items. Such measures include the how-many task (stating a set's total in response to the question, “How many?”) and the give-*n* task (creating a set from a larger set to comply with a verbal request such as “Give me three chips”). Children typically develop both competencies for small numbers (i.e., sets less than four) before they can use a counting procedure to either label a set's total or create a specified set. Small-*n* recognition—children's initial means of stating a set's total number of items—likely involves subitizing, originally defined by Kaufman et al. (1949) as immediately recognizing the total number of items in a set and associating it with an appropriate number word. Small-*n* creation—children's initial means of creating a collection of a verbally specified size—likely also entails using subitizing to identify when a requested number of items have been put out. There is considerable, but not unanimous, agreement that this first subitizing-based phase of cardinality development provides a foundation for the second counting-based phase (Baroody et al., 2006, 2017; Benoit et al., 2004; Carey & Barner, 2019; Fischer, 1992; Klahr & Wallace, 1976; von Glasersfeld, 1982; but cf. Cordes & Gelman, 2005; Gallistel & Gelman, 2000; Nieder, 2017). What is unclear is whether small-*n* recognition and creation emerge simultaneously or in



succession during pre-counting phase. A secondary analysis of preschool pre-counting data (Mix et al., 2012) provided an opportunity to directly test this question, which has significant theoretical and methodological implications.

Theoretical Background

On many accounts, small- n recognition and small- n creation are assumed to emerge simultaneously. In their seminal research on number development, Schaeffer et al. (1974) found that children who could subitize collections of two and perhaps as many as four could, for example, also create sets of two or three objects upon request without counting. They concluded that small- n recognition allows children to create sets of 2 or 3 but did not specify whether the latter emerges simultaneously with or later than the former. In effect, Schaeffer et al. did not clearly specify whether small- n recognition was a necessary and sufficient condition or a necessary condition for small- n creation.

More recently, theorists have argued that small- n recognition and creation unfold simultaneously in a stepwise manner based on the order of magnitude—commonly called n -knower levels (Condry & Spelke, 2008; Le Corre & Carey, 2007, 2008; Le Corre et al., 2006; Sarnecka & Carey, 2008; Wynn, 1990, 1992). Specifically, small- n recognition entails constructing an exact cardinal representation of small numbers by associating a category of sets with number words (e.g., any *pair* of items can be labeled “two”). Initially, such a representation may be inexact, and subitizing a total number may be unreliable (e.g., “two” may simply be understood as “many” or “not ‘one’” and overapplied). However, as the limits of a number-word category are constructed, subitizing becomes a reliable tool for labeling the total of a small set. Although reliable subitizing of “two” may develop simultaneously with that of “one” (Palmer & Baroody, 2011) or even before that of “one” (Beilin, 1975; Durkin et al., 1986; Mix, 2009; Wagner & Walters, 1982), small- n recognition is generally thought to occur first for “one,” then for “two,” next for “three,” and later for somewhat larger numbers. As subitizing skill expands, it permits creating successively larger sets. A “2-knower,” for example, is a child who can reliably recognize or create sets of one and two but not three. (A “subset knower” is a child who is a 1-, 2-, or 3-knower.) On this view, there is no reason to predict that—within each set size or knower-level—performance should differ on small- n recognition and creation tasks.¹

Evidence indicating concordant development of small- n recognition and creation has been taken as support of the discontinuity hypothesis over the continuity hypothesis (Sarnecka & Carey, 2008; Wynn, 1990, 1992)—that an understanding of verbal-based cardinality unfolds in a series of conceptual steps rather than building on an innate understanding of cardinal numbers and counting. As Le Corre et al. (2006, p. 148) noted:

“Such consistency supports the discontinuity hypothesis; it suggests that ‘one’-knowers truly only know the numerical meaning of ‘one,’ ‘two’-knowers truly only know the numerical meaning of ‘one’ and ‘two,’ and so on, across tasks with distinct processing demands ... Finding that children’s [give- n] knower-level is the same as their [how-many] knower-level ... would provide strong evidence in favor of the discontinuity hypothesis.”

Supporters of the continuity hypothesis point to evidence that the give- n task is relatively difficult even when it involves small numbers. For instance, Cordes and Gelman (2005, p. 129) noted that a “child has to create a set of objects, one by one, until she has created a set whose numerical value corresponds to one memory” and that the “combined competence requirements exceed those of a beginning language user.”

A Different Version of the Discontinuity Hypothesis

Addressed first is an argument for a possible succession of small- n skills in the preverbal phase and then some of the methodological implications of this view.

¹ In contrast, it is widely agreed that performance on the how-many and give- n tasks for large sets (> 4) depends on the new and powerful tool of counting and that the development of the former task precedes the latter. This developmental difference has been explained in terms of the additional performance factors required to count out a specified number of items (Resnick & Ford, 1981) or its new conceptual demands (Fuson, 1988; but cf. Sarnecka & Carey, 2008; see Baroody & Lai, 2022, for a review).

Theoretical Argument

Whether the how-many and give- n tasks yield concordant results is not a critical test of the discontinuity (or continuity) hypothesis, and equating non-concordant results with support for the continuity hypothesis and concordant results with support for the discontinuity hypothesis is a false dichotomy. Cordes and Gelman (2005) may have overstated the difficulty of successfully creating a small set, because they assumed subitizing is not a real phenomenon and that children must count out a requested number of items. They were correct, however, that proponents of the discontinuity hypothesis overlook how the give- n task may be more challenging than the how-many task even with small numbers. In the present study, we tested a version of the discontinuity hypothesis that subitizing-based small- n recognition and creation may develop in a non-concordant fashion.

Specifically, although small- n recognition and creation may have a common conceptual basis (verbal-based cardinal concepts of these numbers), performance and conceptual differences in task demands may cause successful small- n creation to emerge later than small- n recognition (if only briefly)—and the latter difference may justify viewing small- n recognition as a necessary (but not a sufficient) condition for small- n creation. These differential task demands include:

1. Small- n creation (the give- n task) encompasses performance factors not required by number-recognition (e.g., the how-many task). Specifically, successively putting out items requires a child to register the requested number in working memory, put out items one at a time, subitize the amount put out, and mentally compare the results of subitizing with the remembered number—a process that needs to be repeated for “give 2” and again for “give 3.” Although putting out the requested number of items simultaneously puts fewer demands on working memory demands, it does require a child to isolate the subset within the larger set and carefully remove only the subset.
2. Conceptually, Benoit et al. (2013) made the theoretical distinction between mapping from a set to a number word (set-to-word mapping) and the reverse (word-to-set mapping).² Small- n recognition entails set-to-word mapping—starting with a specific example of a number and relating it to a number word and its associated general concept (e.g., relating ■■ or 2 to the number word “two” and a cardinal concept of two: any pair of like items, any duality, or even more exactly as one more than one). In contrast, small- n creation requires a word-to-set mapping—starting with a number word and its associated general concept and creating a specific example of the number. These differences in mapping may mean that reliable small- n recognition and creation have different conceptual demands. For example, the how-many task may entail understanding that a set of items can viewed as a total (whole) as well as individual elements (parts), whereas the give- n task requires applying this knowledge—understanding that a set of the specified total needs to be created. Indeed, reliably subitizing small numbers (e.g., accurately, consistently, and selectively labeling sets of 3 as “three”)—a set-to-word mapping—seems necessary for creating a requested number of items—a word-to-set mapping—via subitizing.

Based on these considerations and others detailed in Baroody et al. (2017) and Baroody and Lai (2022), we propose an alternative course of verbal cardinality development summarized in Table 1. This model diverges from the conventional wisdom regarding Phase 1 in that it posits successive development of small- n recognition and creation—a proposition that is tested by the present analysis. For evidence supporting the sequential development of analogous subphases for Phase 2, see Baroody et al. (2022) and Baroody and Lai (2022).

² Benoit et al. (2013) used a how-many task to gauge set-to-word mapping and a task that involved matching a spoken number to one of six arrays of dots to assess word-to-set mapping. They found that 3-year-olds did not significantly differ on the two tasks for sets of one to three and concluded that small- n set-to-word and word-to-set mappings were equally difficult. However, Benoit et al.'s participants performed noticeably better on the set-to-word (how-many) task ($M = 5.06$ correct of 6 possible trials involving 1 to 3) than on the word-to-set (forced-choice matching) task ($M = 4.06$ of 6 possible trials involving 1 to 3). This difference may not have reached statistical significance because (a) collapsing data over arrays of one to three may have masked a real difference for sets of three or perhaps even sets of two and (b) the relatively small sample ($n = 16$) provided insufficient power to detect a real difference. Moreover, a possible confound was that the two types of mapping were assessed by different types of tasks. The set-to-word (how-many) measure involved a production task, whereas the word-to-set mapping measure entailed a possibly easier forced-choice matching task. The issue of simultaneous or successive development of the two mappings with small numbers needs to be re-evaluated with (a) a sample that would provide sufficient power to detect real differences; (b) separate data analyses for set sizes 1, 2, and 3; and (c) analogous tasks for each mapping. Regarding the last point, both answer-production tasks such as the how-many task for set-to-word mapping and the give- n task for word-to-set mapping and forced-choice matching tasks for both types of mappings could be used.

Table 1

Possible Phases of Verbal-Based Cardinality Development, Their Conceptual Basis, Type of Mapping, and Measures (Baroody & Lai, 2022; Baroody et al., 2017)

Aspect of Cardinal Number	Conceptual Basis	Mapping	Direct Measure
First (Pre-Counting) Phase of Cardinality Development—Before Meaningful 1-1 Counting (i.e., before understanding of the CP)			
Subphase 1.1. Small- <i>n</i> recognition: subitizing-based number recognition (commonly called <i>n</i> -knower levels)	An exact <i>cardinal representation</i> of a small number underlies the reliable ability to subitize (immediately recognize) 1, 2, or 3	Set-to-Word (via subitizing)	How-many task without counting
Subphase 1.2. Small- <i>n</i> creation: subitizing-based set creation (also commonly called <i>n</i> -knower levels)	An exact <i>cardinal representation</i> of small numbers can be used to subitize when 1, 2 or 3 have been put out (i.e., to reliably stop the set-creation process)	Word-to-Set (via subitizing)	Give- <i>n</i> task without counting
Second (Counting) Phase of Cardinality Development—Meaningful 1-1 Counting (i.e., with understanding of the CP)			
Subphase 2.1. Counting-based number identification (commonly called CP-knower level)	<i>Cardinality Principle (CP)</i> or what Fuson (1988) calls the <i>count-cardinal concept</i> : the last number word used in counting a collection represents its total number ^a	Set-to-Word (via counting)	Counting-based how-many task
Subphase 2.2. Counting-based number creation— (also commonly called CP-knower level)	What Fuson (1988) calls the <i>cardinal-count concept</i> : a cardinal number would be the last number word if a collection is counted	Word-to-Set (via counting)	Counting-based give- <i>n</i> task

Note. Unlike stages in which a successive stage replaces a prior stage, the second phase of verbal cardinality development does not replace the first but (greatly) supplements it. CP indicates the cardinality principle.

^aFuson (1988) noted that prior to the CP, some children learn to respond to how-many questions with a last-word rule (repeating the last count word by rote—without realizing it represents the total number of items).

Methodological Considerations

According to the model outlined in Table 1, although direct measures of Phase-2 competencies would involve counting, direct measures of Phase-1 competencies would involve subitizing, not counting. For example, for their pioneering research, Schaeffer et al. (1974) assessed the Phase-1 competence of small-*n* recognition with a how-many task that did not involve counting. Participants were shown a pictured array of one to four men and asked, “How many men are there?” If a participant counted, an experimenter requested the child not count or point but simply tell how many men there were. Researchers now often avoid using a how-many task with counting to assess Phase-2 knowledge, because (transitional) children may use the last-word rule learned by rote to be successful, thus overestimating knowledge of the cardinality principle (CP) or achievement of Subphase 2.1 in Table 1 (Sarnecka & Carey, 2008). However, the same concern does not apply to assessing Phase-1 small-*n* recognition with the how-many task via subitizing (without counting or need to apply the CP).

Some versions of the give-*n* task do not involve counting but some do, and such variations yield different results (Sella et al., 2021). Krajcsi (2021) found that prompting counting on the give-*n* task can minimize performance errors when assessing Phase-2 knowledge. Specifically, among CP knowers, this version of the give-*n* task indeed resulted in more success than not prompting counting. However, prompting counting on the give-*n* task when assessing Phase-1 pre-counting small-*n* creation skill may not be helpful and may even be counterproductive. Asking Phase-1 children to use developmentally more advanced Phase-2 concepts and skills to check small-*n* creation efforts is likely to be incomprehensible or even confusing, be interpreted as challenging a child’s initial response, and promote disengagement from the task—all of which may lead to underestimating competence of small-*n* creation ability.

Although the previously stated conjecture needs systematic examination, two recent findings are consistent with this proposition. Krajcsi (2021) concluded that a counting follow-up did not benefit subset-knowers (Phase-1 children).

Marchand et al. (2022) used two versions of a give- n task with a counting prompt and found for both the reliability “of individual knower levels varied considerably, such that non-knowers, 1-knowers, 2-knowers, and CP-knowers exhibited fairly high [reliability], while 3-, 4-, and 5-knowers did not” (p. 12).

Empirical Evidence Regarding Developmental Order

The surprisingly few comparisons of young children’s small- n recognition and creation provide no clear evidence about their developmental relation. Schaeffer et al.’s (1974) data are not presented in sufficient detail to determine whether small- n recognition and creation emerged in tandem or sequentially. More recently, Mou et al. (2021) used how-many and give- n tasks that did not instruct children to count initially or follow-up with a request to check via counting. Using latent modeling of 3- and 4-year-olds’ performance on the how-many and give- n tasks with sets of up to eight items, they 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 how-many and give- n tasks gauge interchangeable concepts and are consistent with multiple dimensions of cardinal-number knowledge acquisition.

Neither the Schaeffer et al. (1974) nor the Mou et al.’s (2021) study addressed the discontinuity hypothesis we have proposed because data were not analyzed separately for each small number. The latter’s analysis also included data for both the how-many and give- n tasks beyond the subitizing range (i.e., required counting to quantify). Moreover, as Mou et al.’s participants included 4-year-olds and had a mean age of 3 years and 11 months, it seems likely that the vast majority exhibited (near) ceiling performance for small- n recognition and creation. Finally, and most importantly, Mou et al.’s “how-many” task did not involve a how-many question, and success was defined as counting a set correctly—despite research that indicates children can accurately count one-to-one before accurately labeling the cardinality of sets (Schaeffer et al., 1974).

Wynn (1990, p. 155) concluded that a comparison of twenty-four 2- and 3-year-olds’ “performance across the ‘how-many’ and ‘give-a-number’ tasks shows strong within-child consistency” regarding when the CP develops. Because she focused on when the CP (Subphase 2.1 in Table 1) emerges, children were asked to count even on small- n trials of both the how-many and give- n tasks. With the latter task, this occurred if a child did not spontaneously count, whether the initial response was correct or not. Wynn’s (1990, 1992) results, then, do not bear on developmental relation between small- n recognition and creation in Phase 1.

Le Corre et al. (2006) used the “what’s on this card” (WOC) task to assess small- n recognition and Wynn’s (1990, 1992) give- n task with a counting follow-up to gauge small- n creation. They concluded from their analysis using the Wilcoxon Signed-Ranks test:

“The two tasks were highly consistent ... While more children had higher knower-levels on WOC than [give- n] ($n = 12$) than the other way around ($n = 5$), this was not significant, $Z = 0.79$, $p = 0.4$. Thus, there was no evidence that children’s knower-levels were systematically higher on WOC than on give- n ” (p. 150).

The non-significant result, though, may simply have been due to a lack of power. As ties (33 of their 50 cases) are not considered in the Wilcoxon test, the actual or redefined n was only 17. A power analysis using G*Power indicated the probability of correctly rejecting the null hypothesis was either 0.22 (one-tailed) or 0.13 (two-tailed). So inversely, the probability of Type II error would have been .78 and .87, respectively. Moreover, if give- n 0- and 1-knower levels listed in Le Corre et al.’s Table 3 are combined into a single category, only 58% (18 of 31) of the n -knowers produced concordant results. Even differences of one level represent an appreciable difference in the estimation a child’s conceptual understanding of small numbers.

Rationale for the Present Study

A post-hoc analysis of the Mix et al. (2012) intervention study provided the first opportunity to directly address the issue of whether small- n recognition and creation develop simultaneously or successively using tasks that do not involve

counting (i.e., confound Phase-1 and Phase-2 competencies) and analyzing the data of each small number separately. Specifically, the analysis compared 3-year-olds' performance on how-many and give-*n* tasks that disallowed one-to-one counting and did so separately for sets of one, two, and three.

Testing at one time point may miss the transition from small-*n* recognition (possible Subphase 1.1 in Table 1) to small-*n* creation (possible Subphase 1.2 in Table 1), especially if two subphases develop in rapid succession. Put differently, a one-shot assessment is more likely than multiple assessments to test children before achieving either subphase or after achieving both subphases. For this reason, children were tested three times on each task to check for possible transitions that indicate prior success on small-*n* recognition (or vice versa).

Method

The intervention study reported by Mix et al. (2012) focused on different methods of modeling the CP and provided no training on either small-*n* recognition or creation.

Participants

The Mix et al. (2012) study involved 60 participants ($M = 3$ years; 7 months, $SD = 0;3$, range = 3;1–4;7) recruited from preschool programs serving predominantly Caucasian middle-class communities in two small cities in Indiana and Michigan. Informed consent was obtained for experimentation with human subjects.

Procedure

Children were tested three times about 3 weeks apart—originally intended as the pretest (Time 1 or T1), immediate posttest (T2), and delayed posttest (T3).

Tasks

For each trial of the how-many task, children were asked to tell how many objects were displayed on a 5- × 8-inch index card. For each of the three time points, there were two cards for each collection size 1 to 3, resulting in a total of six trials per number. Each set consisted of identical, photographic images arranged in a random array. On each trial, the experimenter held up a card and asked, “How many are there?” As 3-year-olds typically do not count unless they can touch the objects with a finger, the cards were held out of a child's reach to eliminate counting. If a child was close enough to a card to touch it, the tester pulled the card out of reach. No feedback was provided. A response to the how-many trial of two and three was scored as correct if a child correctly indicated the cardinal value of a collection without behaviors consistent with one-to-one counting, namely counting from “one” to the cardinal value (e.g., for three items, counting: “One, two, three”) or successive pointing to each item. It is not possible to distinguish between subitizing and counting with collections of one. Therefore, these trials were scored as correct if a child labeled a single item “one” whether the child pointed.

For the give-*n* task, children were given a pile of 15 objects and asked to create a set (e.g., “Give me three pigs”). Sets of one, two, and three were each requested twice in a random order that was interspersed with requests for five and six. Participants were not given instructions on counting, because Mix et al. (2012) allowed children to choose a strategy that was appropriate for either small numbers or larger ones: “Although children can produce small sets on demand without understanding [the CP], the ability to produce sets greater than 4 is taken as evidence for [CP] understanding because these larger sets must be counted (i.e., they cannot be subitized) (Wynn, 1990)” (p. 277). As children tend to choose a strategy that ensures success but requires the least effort, the assumption that young children would use subitizing on sets of 1 to 3 on the give-*n* task seems reasonable. However, even if a child had to rely on counting for success with small numbers, this would work against our hypothesis that successful performance on the how-many task emerges before success on the give-*n* task.

For purposes of the present re-analysis, performance on only small-*n* sets was considered. A child was scored as correct on a trial if the number created matched the number requested. Unlike typical *n*-knower scoring, producing a

non-requested number for a trial was not penalized (e.g., producing three for a give-2 trial did not count against a child's give-3 total score). However, overestimating “knower level” works against the authors' hypothesis that the ability to recognize three precedes the ability to produce three.

Analyses

A Kolmogorov-Smirnov test confirmed that the data were not normally distributed. Thus, a non-parametric regression test was used to check whether session (T1, T2, T3), task (how-many vs. give- n), and set size ($n = 1, 2, 3$ using 2 as the reference set) had a significant impact on the outcome or dependent variable (number correct: 0, 1, or 2). Specifically, a proportional odds model was used because the dependent variable was ordered into three categories. Let Y be an ordinal outcome with J categories; the model can be defined as:

$$\text{logit}(P(Y \leq j)) = \log\left(\frac{P(Y \leq j)}{P(Y > j)}\right) = \beta_{j0} + \beta_{j1}x_1 + \dots + \beta_{jp}x_p,$$

where $\beta_{j0} + \beta_{j1}x_1 + \dots + \beta_{jp}x_p$ are model coefficient parameters (i.e., intercepts and slopes), with p predictors for $j = 1, 2, \dots, J - 1$. Intercepts can differ, but slopes are constant across categories due to the proportional odds assumption. Hence, the proportional odds model can be simplified as:

$$\text{logit}(P(Y \leq j)) = \log\left(\frac{P(Y \leq j)}{P(Y > j)}\right) = \beta_{j0} + \beta_1x_1 + \dots + \beta_px_p.$$

As Dixon and Moore (2000) argued that it is not enough to corroborate a hypothesis of developmental order but that alternative hypotheses need to be disconfirmed, a follow-up analysis involved comparing three possible developmental hypotheses: (a) synchronous-development hypothesis (simultaneous development of how-many and give- n competence), (b) how-many-priority hypothesis (earlier development of how-many competence), and (c) give- n -priority hypothesis (earlier development of give- n competence). In a 3 x 3 table, perfect support for the how-many-priority hypothesis over the simultaneous-development hypothesis would occur if all the data were distributed in three cells: partially successful on the how-many task but unsuccessful on the give- n task and successful on the how-many task but partially successful or unsuccessful on the give- n task (Dixon & Moore). Note that two cells (completely successful on both tasks and unsuccessful on both tasks) are consistent with all three hypotheses and, thus, not useful in discerning developmental order.

Results

The number correct by time, task, and set size are summarized in Figure 1. The non-parametric regression analysis was performed with R software. The proportional odds assumption was checked to see if it held. As Table 2 indicates, the test was insignificant. As the null hypothesis cannot be rejected (i.e., the proportional odds assumption holds), the proportional odds model is suitable for the data.

Table 3 shows the model estimates. All variables are significant except for time. Results of the odds ratios and the confidence intervals in Table 4 confirm that time had no effect on the response of students ($OR = 0.94$; 95% CI [0.67, 1.3]; $OR = 1.06$; 95% CI [0.75, 1.50]). As indicated by the odds ratio of 1.952 (95% CI [1.47, 2.60]), the odds of getting a higher score on the how-many task than on the give- n task (e.g., 2 or 1 on how many versus 0 on give n) are almost twice that of the reverse, holding all other variables constant. An OR of 1.68, 3.47, and 6.71 are equivalent to a small, medium, and large effect size (Cohen's d), respectively (Chen, Cohen, & Chen, 2010). Holding all other variables constant, the odds of getting a higher score on Set Size 1 are about three times greater than on Set Size 2 ($OR = 3.165$; 95% CI [0.21, 0.48]); the odds of getting a higher score on Set Size 2 are about 0.44 times greater than on Set Size 3 ($OR = 0.443$; 95% CI [0.32, 0.59]). The interaction between task and set size was not significant.

Figure 1

Students' Scores by Task, Set Size, and Session

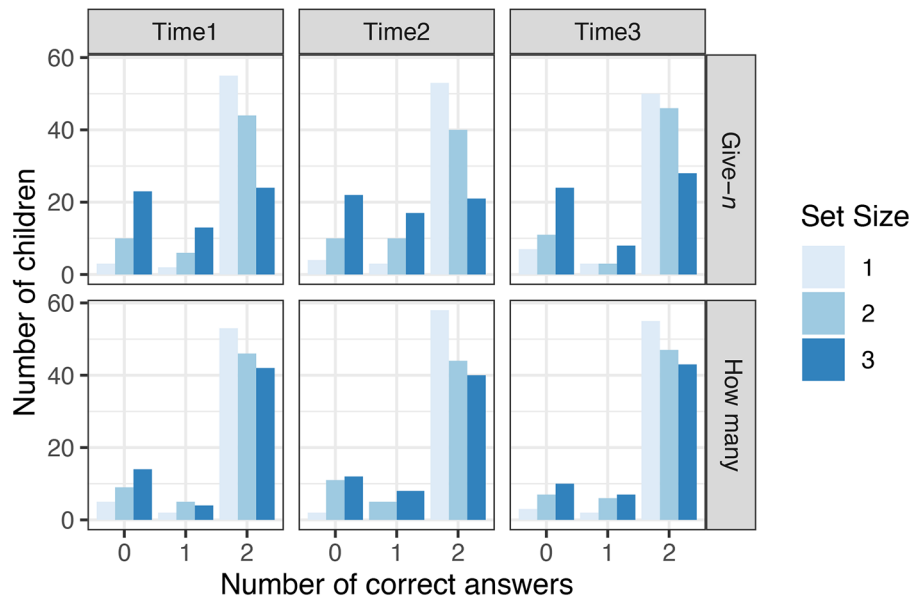


Table 2

Test Results for the Proportional Odds Assumption

Test for	χ^2	df	probability
Omnibus	7.28	5	0.20
Set Size 2	1.08	1	0.30
Set Size 3	3.16	1	0.08
Task – How Many	1.54	1	0.21
Time 2	1.69	1	0.19
Time 3	0.09	1	0.76

Table 3

Summary of the Proportional Odds Model

Variable	Value	SE	t	p
Set Size 1	-1.152	0.214	-5.391	< .001
Set Size 3	-0.837	0.158	-5.279	< .001
Task – How Many	0.669	0.145	4.617	< .001
Time 2	-0.066	0.173	-0.384	.701
Time 3	0.058	0.177	0.328	.743
0 1	-1.381	0.174	-7.938	< .001
1 2	-0.747	0.169	-4.434	< .001

Table 4

The Odds Ratios and Confidence Intervals

Variable	OR	2.50%	97.50%
Set Size 1	3.165	2.099	4.863
Set Size 3	0.443	0.317	0.589
Task – How Many	1.952	1.472	2.598
Time 2	0.936	0.666	1.313
Time 3	1.060	0.749	1.500

Table 5

Number of Correct Responses on the How-Many and Give-n Tasks by Set Size and Time Point

Set size ↓	Number correct ↓ →	Session 1			Session 2			Session 3						
		Give-n task			Give-n task			Give-n task						
		0	1	2	0	1	2	0	1	2				
N = 1	How-	2	A 3	B 2	C 48	A 4	B 2	C 52	A 5	B 3	C 47			
	many	1	D 0	E 0	F 2	D 0	E 0	F 0	D 1	E 0	F 1			
	task	0	G 0	H 0	I 5	G 0	H 1	I 1	G 1	H 0	I 2			
			A-B-D v E 2-tailed sig. Sign Test			0.0625			0.0313*			0.0039**		
			F-H-I v E 2-tailed sig. Sign Test			0.0156			0.5000			0.2500		
			A-B-D v F-H-I 2-tailed sig. Sign Test			0.7744			0.2891			0.1460		
N = 2	How-	2	A 2	B 4	C 40	A 2	B 6	C 36	A 4	B 1	C 42			
	many	1	D 1	E 2	F 2	D 2	E 1	F 2	D 3	E 1	F 2			
	task	0	G 7	H 0	I 2	G 6	H 3	I 2	G 4	H 1	I 2			
			A-B-D v E 2-tailed sig. Sign Test			0.1797			0.0117*			0.0391*		
			F-H-I v E 2-tailed sig. Sign Test			0.6875			0.0703			0.2188		
			A-B-D v F-H-I 2-tailed sig. Sign Test			0.5488			0.6291			0.5811		
N = 3	How-	2	A 9	B 10	C 23	A 8	B 12	C 20	A 11	B 5	C 27			
	many	1	D 3	E 1	F 0	D 3	E 5	F 0	D 5	E 1	F 1			
	task	0	G 11	H 2	I 1	G 11	H 0	I 1	G 8	H 2	I 0			
			A-B-D v E 2-tailed sig. Sign Test			< 0.0001***			0.0009***			< 0.0001***		
			F-H-I v E 2-tailed sig. Sign Test			0.2500			0.2188			0.2500		
			A-B-D v F-H-I 2-tailed sig. Sign Test			0.0002***			< 0.0001***			0.0003***		

Note. Excluding Cells C and G, which are consistent with all three hypothesis, the data in unshaded Cell E are consistent with the simultaneous hypothesis (simultaneous development of how-many and give-n competence); that in the green-shaded Cells A, B, and D, with the how-many-priority hypothesis (earlier development of how-many competence); and that in the red shaded Cells F, H, and I, a give-n-priority hypothesis (earlier development of give-n competence).

* $p < .05$. ** $p < .01$. *** $p < .001$.

As the omnibus analysis was significant for task and set size, a follow-up analysis was conducted to examine further the developmental relation between the how-many and give- n tasks by each set size. This analysis was done by time point to maintain independent observations. The participants' performance on the how-many and give- n tasks by collection size and time point are summarized in Table 5. A comparison of the data consistent with the how-many-priority hypothesis indicated by the green-shaded cells in Table 5—Cell A (successful on the how-many task but unsuccessful on the give- n task), Cell B (successful on the how-many task but partially successful on the give- n task), and Cell D (partially successful on the how-many task but unsuccessful on the give- n task) and that consistent with the synchronous-development hypothesis (Cell E; partially successful on both tasks)—revealed a significant difference in favor of the former hypothesis in seven of the nine cases. For the set size of three, the how-many-priority hypothesis was significantly superior to both the simultaneous-development hypothesis *and* the give- n -priority hypothesis (the data in the red-shaded Cells F, H, and I).

Discussion

The results of the omnibus analysis indicate that performance on each task was relatively stable over the three testing sessions, significantly higher on the how-many task than on the give- n task, and significantly different by set size ($1 > 2$ and $2 > 3$). As the follow-up analysis clarifies, the omnibus analysis does not support a strong version of the how-many-priority hypothesis—that children succeed on the how-many task with 1, 2, and 3 before they do so on the give- n task with 1, 2, and 3. Instead, consistent with authors' alternative discontinuity view and contrary to the conventional wisdom (simultaneous-development hypothesis), the follow-up analysis generally supported a weak version of the how-many first hypothesis. Specifically, it indicated that, for sets of 1 and 2, prior success on the how-many task generally occurred significantly more often than simultaneous success on both tasks but not significantly more often than prior success on the give- n task. In contrast to the inconclusive results for sets of 1 and 2, those for sets of 3 were clearcut—the how-many-priority hypothesis was significantly superior to both the simultaneous and give- n -priority hypotheses.

The lack of conclusive results for sets of one and two is likely due to a ceiling effect—too few non-concordant cases to overcome measurement error. The present results are consistent with Marchand et al.'s (2022) finding of higher reliability for 1- and 2-knowers than for 3-knowers. Children often construct verbal-based number concepts in a step-like fashion—an understanding of “one,” then “two,” and finally “three” or, in some cases, “one” and “two” before “three” (Mix, 2009; Palmer & Baroody, 2011). As most participants were 3.5-years of age or older and children this age can typically recognize and create sets of one and two, it makes sense that at least 60% of the participants in the present study were successful on both tasks with sets of one and two.

Further research is needed with 2-year-olds—with children who are just constructing verbally based concepts of “one” and “two”—to evaluate whether competence with the how-many task emerges simultaneously or successively with that for the give- n task for sets of one and two. In brief, although further research with younger and less developmentally advanced children is needed, it should not be taken for granted that how-many and give- n tasks will yield the equivalent results with small collections (e.g., knower levels), particularly those involving three items.

It could be argued that the scoring procedure of the give- n task used in the present research—unlike that for Wynn's (1990, 1992) give- n task—did not check for overapplication of a number word and, thus, overestimated small- n creation competence. However, ignoring such possible overapplications is not a threat to internal validity. Overestimating give- n competence works against the omnibus finding that performance on the how-many task was significantly greater than that on the give- n task or the follow-up analysis supporting the how-many-priority hypothesis over the simultaneous-development hypothesis for all small sets and over the give- n -priority hypothesis for sets of three. However, not checking for overapplications on the give- n task does limit the external validity of the present results. That is, caution should be exercised in generalizing these results to cases that involved checking for overapplications. Moreover, if a give- n task is needed to *accurately* gauge, for example, a child's n -creator level, scoring should account number-word overapplications.

Implications and Conclusions

Theoretical Implications

Researchers have focused on whether performance on small- n recognition and creation tasks are concordant because such results were interpreted as supporting the discontinuity hypothesis (e.g., Le Corre et al., 2006), whereas non-concordant results were regarded as support for the continuity hypothesis (e.g., Cordes & Gelman, 2005). The present results are a first step toward supporting a version of the discontinuity hypothesis that entails postulating non-concordant development of subitizing-based small- n recognition and creation.

Marchand et al. (2022) offered two reasons for the instability of higher subset levels: (a) misclassification of CP-knowers and (b) noisy associative mappings between number words and approximate magnitudes (see also Krajcsi & Fintor, 2022; Wagner & Johnson, 2011). A third reason—children’s progressive construction of verbal-based number concepts—could explain the present, inconclusive results with sets of one and two due to a ceiling effect but clearcut results with sets of three and could either work in tandem with the second reason just discussed or not. Like other verbal-based concepts, children may initially overgeneralize the associated word and only gradually apply the word accurately and reliably (Mix, 2009; Palmer & Baroody, 2011). If our alternative discontinuity hypothesis outlined in Table 1 is correct and a child has already constructed exact verbal concepts for “one” and “two” but not for “three,” then significant non-concordant results can be expected only between the recognition of three and the creation of three—whether exact verbal small-number concepts build on an approximate-number system. Specifically, if children have an inexact concept of “three” as “many,” a fragile concept of “three,” or a newly emerged exact concept of “three,” then there is a greater chance they will perform (more) successfully on the recognition-of-three task than on the create-three task, whether associations between number words and the approximate-number system are a factor.

If further research confirms that small- n recognition emerges before small- n creation for some or all three of the smallest whole numbers (i.e., corroborate that Subphase 1.1 and Subphase 1.2 in Table 1 are distinct), it would be inappropriate to refer to both competencies as n -knower levels. More accurate labels for these competencies might be the “ n -recognizer levels” and “ n -creator levels,” respectively (cf. Clements & Sarama, 2021). Another reason for using the more specific terms n -recognizer and n -creator levels (instead of the broader term n -knower levels) was adduced by Barner and Bachrach (2010). They observed that specifying a particular n -knower level could be misleading, because it implies that a child does not have knowledge of numbers beyond the level. Their evidence and that of others (Gunderson et al., 2015; Krajcsi & Fintor, 2022; O’Rear et al., 2020; Sarnecka & Gelman, 2004; Wagner et al., 2019) indicates that children have some understanding of numbers beyond their n -knower level (e.g., knowledge of approximate magnitude).

Methodological Implications

As indicated in Table 1, caution should be exercised if the give- n task *without* counting is used to gauge the *first phase of cardinality knowledge generally* (i.e., n -knower knowledge that encompasses both small- n recognition or Sublevel 1.1 and small- n creation or Sublevel 1.2). The present results indicate that this task may underestimate the three-recognizer step of Sublevel 1.1 with 3-year-olds. Furthermore, in an intensive and dense case study of a toddler from 18 to 49 months of age, Palmer and Baroody (2011) found that, at 29 months, the child had difficulty responding to requests of “give me two” even after achieving reliable identification of sets of two. Further research is needed to examine whether the give- n task may underestimate the two-recognizer (or even one-recognizer) step of Sublevel 1.1 with 2-year-olds. The give- n task without counting is useful IF the goal is a *conservative* estimate of 3-year-olds’ Phase-1 cardinality of knowledge of three (or possibly two or even one if testing 2-year-olds).

The common practice in cognitive, developmental, and educational psychology of using the give- n task *with* counting to assess subset knowers needs careful reconsideration. For example, for Wynn’s (1990) version of the task, “any child who did not spontaneously count the objects was prompted to count ... (e.g., “Can you count and make sure there are two?”; p. 171). However, repeatedly challenging children who have not constructed the CP and who do not understand the purpose of one-to-one counting to check their initial subitizing-based effort by counting could be viewed as challenging their initial answers and undermine confidence in them. Although research is needed to confirm the implication, asking subset knowers to count may be confusing to them, may render the task more taxing for no

apparent reason, and may result in underestimating competence because of disinterest (avoidance behaviors) or acting out (uncooperative behaviors).

Funding: Preparation of this report was supported by the Institute of Education Science [grant number R305A150243] and the National Science Foundation [grant numbers 1621470 & 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.

Acknowledgments: The authors thank two anonymous reviewers for their most helpful comments in writing the report.

Competing Interests: The authors have declared that no competing interests exist.

CRedit Author Statement: **First Author:** Conceptualization, Formal analysis—Follow-up analysis, Writing—Original draft, Writing—Review & Editing; **Second Author:** Formal analysis—Power analysis, Writing—Review & Editing; **Third Author:** Formal analysis—Nonparametric statistical analyses; **Fourth Author:** Formal analysis—Kolmogorov-Smirnov analysis, Writing—Review & Editing.

Data Availability: The data used for the present report are freely available in the [Supplementary Materials](#) or by e-mailing the first author <baroody@illinois.edu>.

Supplementary Materials

The Supplementary Materials contain the research data for this study (for access see [Index of Supplementary Materials](#) below).

Index of Supplementary Materials

Baroody, A. J., Mix, K. S., Kartal, G., & Lai, M.-I. (2023). *Supplementary materials to "The development and assessment of early cardinal-number concepts"* [Research data]. PsychOpen GOLD. <https://doi.org/10.23668/psycharchives.12521>

References

- Barner, D., & Bachrach, A. (2010). Inference and exact numerical representation in early language development. *Cognitive Psychology*, 60(1), 40–62. <https://doi.org/10.1016/j.cogpsych.2009.06.002>
- Baroody, A. J., Clements, D. H., & Sarama, J. (2022). Lessons learned from 10 experiments that tested the efficacy and assumptions of hypothetical learning trajectories. *Education in Science*, 12, Article 195. <https://doi.org/10.3390/educsci12030195>
- Baroody, A. J., & Lai, M.-I. (2022). The development and assessment of counting-based cardinal-number concepts. *Educational Studies in Mathematics*, 111(2), 185–205. <https://doi.org/10.1007/s10649-022-10153-5>
- Baroody, A. J., Lai, M.-I., & 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., Lai, M.-I., & Mix, K. S. (2017, October). Assessing early cardinal-number concepts. In *Proceedings of the Thirty-ninth Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (p. 324). Indianapolis, IN, USA.
- Beilin, H. (1975). *Studies in the cognitive basis of language development*. Academic Press.
- Benoit, L., Lehalle, H., & Jouen, F. (2004). Do young children acquire number words through subitizing or counting? *Cognitive Development*, 19(3), 291–307. <https://doi.org/10.1016/j.cogdev.2004.03.005>
- Benoit, L., Lehalle, H., Molina, M., Tijus, C., & Jouen, F. (2013). Young children's mapping between arrays, number words, and digits. *Cognition*, 129(1), 95–101. <https://doi.org/10.1016/j.cognition.2013.06.005>
- 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>
- Chen, H., Cohen, P., & Chen, S. (2010). How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Communications in Statistics – Simulation and Computation*, 39(4), 860–864. <https://doi.org/10.1080/03610911003650383>

- Clements, D. H., & Sarama, J. (2021). *Learning and teaching early math: The learning trajectories approach* (3rd ed). Routledge.
- 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*(1), 22–38. <https://doi.org/10.1037/0096-3445.137.1.22>
- Cordes, S., & Gelman, R. (2005). The young numerical mind: When does it count? In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 127–142). Psychology Press.
- Dixon, J. A., & Moore, C. F. (2000). The logic of interpreting evidence of developmental ordering: Strong inference and categorical measures. *Developmental Psychology*, *36*(6), 826–834. <https://doi.org/10.1037/0012-1649.36.6.826>
- Durkin, K., Shire, B., Riem, R., Crowther, R. D., & Rutter, D. R. (1986). The social and linguistic context of early number word use. *British Journal of Developmental Psychology*, *4*(3), 269–288. <https://doi.org/10.1111/j.2044-835X.1986.tb01018.x>
- 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.
- Fuson, K. C. (1988). *Children's counting and concepts of number*. Springer.
- 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)
- Gunderson, E. A., Spaepen, E., & Levine, S. C. (2015). Approximate number word knowledge before the cardinal principle. *Journal of Experimental Child Psychology*, *130*, 35–55. <https://doi.org/10.1016/j.jecp.2014.09.008>
- Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkman, J. (1949). The discrimination of visual number. *The American Journal of Psychology*, *62*(4), 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*, Article 100968. <https://doi.org/10.1016/j.cogdev.2020.100968>
- Krajcsi, A., & Fintor, E. (2022). A refined description of initial symbolic number acquisition. *Cognitive Development*, *62*(4), Article 101288. <https://doi.org/10.1016/j.cogdev.2022.101288>
- 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*(2), 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*(2), 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*(2), 130–169. <https://doi.org/10.1016/j.cogpsych.2005.07.002>
- Marchand, E., Lovelett, J. T., Kendro, K., & Barner, D. (2022). Assessing the knower-level framework: How reliable is the Give-a-number task? *Cognition*, *222*(4), Article 104998. <https://doi.org/10.1016/j.cognition.2021.104998>
- Mix, K. S. (2009). How Spencer made number: First uses of the number words. *Journal of Experimental Child Psychology*, *102*(4), 427–444. <https://doi.org/10.1016/j.jecp.2008.11.003>
- Mix, K. S., Sandhofer, C. M., Moore, J. A., & Russell, C. (2012). Acquisition of the cardinal word principle: The role of input. *Early Childhood Research Quarterly*, *27*(2), 274–283. <https://doi.org/10.1016/j.ecresq.2011.10.003>
- 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. <https://doi.org/10.1016/j.tics.2017.03.014>
- O'Rear, C. D., McNeil, N. M., & Kirkland, P. K. (2020). Partial knowledge in the development of number word understanding. *Developmental Science*, *23*(5), Article e12944. <https://doi.org/10.1111/desc.12944>
- Palmer, A., & Baroody, A. J. (2011). Blake's development of the number words “one,” “two,” and “three”. *Cognition and Instruction*, *29*(3), 265–296. <https://doi.org/10.1080/07370008.2011.583370>
- Resnick, L. B., & Ford, W. W. (1981). *The psychology of mathematics for instruction*. Erlbaum.
- Sarnecka, B. W., & Carey, S. (2008). How counting represents number: What children must learn and when they learn it. *Cognition*, *108*(3), 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*(3), 329–352. <https://doi.org/10.1016/j.cognition.2003.10.001>
- Schaeffer, B., Eggleston, V. H., & Scott, J. L. (1974). Number development in young children. *Cognitive Psychology*, *6*(3), 357–379. [https://doi.org/10.1016/0010-0285\(74\)90017-6](https://doi.org/10.1016/0010-0285(74)90017-6)
- Sella, F., Slusser, E., Odic, D., & Krajcsi, A. (2021). The emergence of children's natural number concepts: Current theoretical challenges. *Child Development Perspectives*, *15*(4), 265–273. <https://doi.org/10.1111/cdep.12428>
- von Glasersfeld, E. (1982). Subitizing: The role of figural patterns in the development of numerical concepts. *Archives de Psychologie*, *50*(194), 191–218.
- Wagner, J. B., & Johnson, S. C. (2011). An association between understanding cardinality and analog magnitude representations in preschoolers. *Cognition*, *119*(1), 10–22. <https://doi.org/10.1016/j.cognition.2010.11.014>
- Wagner, K., Chu, J., & Barner, D. (2019). Do children's number words begin noisy? *Developmental Science*, *22*(1), Article e12752. <https://doi.org/10.1111/desc.12752>
- Wagner, S. H., & Walters, J. A. (1982). A longitudinal analysis of early number concepts. In G. Forman (Ed.), *Action and thought: From sensorimotor schemes to symbolic operations* (pp. 137–161). Academic Press.
- Wynn, K. (1990). Children's understanding of counting. *Cognition*, *36*(2), 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*(2), 220–251. [https://doi.org/10.1016/0010-0285\(92\)90008-P](https://doi.org/10.1016/0010-0285(92)90008-P)



Journal of Numerical Cognition (JNC) is an official journal of the Mathematical Cognition and Learning Society (MCLS).



leibniz-psychology.org

PsychOpen GOLD is a publishing service by Leibniz Institute for Psychology (ZPID), Germany.