University of Connecticut
OpenCommons@UConn

Honors Scholar Theses | Honors Scholar Program

Summer 8-31-2020

Exploring the Relationship Between Children's Vocabulary and their Understanding of Cardinality: A Methodological Approach

Justin Slifer
justin.slifer@uconn.edu

Emily Carrigan
emily.carrigan@uconn.edu

Kristin Walker
kristin.walker@uconn.edu

Marie Coppola
marie.coppola@uconn.edu

Follow this and additional works at: https://opencommons.uconn.edu/srhonors_theses

Part of the Developmental Psychology Commons, and the Statistical Methodology Commons

Recommended Citation
Slifer, Justin; Carrigan, Emily; Walker, Kristin; and Coppola, Marie, "Exploring the Relationship Between Children's Vocabulary and their Understanding of Cardinality: A Methodological Approach" (2020). *Honors Scholar Theses*. 745.
https://opencommons.uconn.edu/srhonors_theses/745

Exploring the Relationship Between Children's Vocabulary and their Understanding of Cardinality: A Methodological Approach

Justin Slifer - slifer.justin@yahoo.com
Thesis Advisor: Marie Coppola
Honors Advisor: Alexander Teplyaev
Other Collaborators: Emily Carrigan, Kristin Walker
Mathematics/Statistics
University of Connecticut

Abstract

Is there a relationship between vocabulary and children's understanding of cardinality? Does the way in which we classify cardinality data as tested by the Give-a-Number task affect finding such a relationship? This thesis explored these questions using a methodological approach, by testing the relationship between children's receptive vocabulary scores and Give-a-Number scores classified in two different ways, the traditional knower-level assessment, as well as by calculating the proportion of trials answered correctly. A significant correlation was found between participants' receptive vocabulary scores and Give-a-Number scores using both manners of classification, independent of the children's ages. The results were compared with those of past studies to examine what discrepancies exist regarding the relationship, and what factors may be contributing to the discrepancies. Though it was not reflected in our dataset, after comparing the results of several studies, we determined that the manner of classification (i.e., knower-level vs. proportion-correct), as well as age, may have an effect on the existence and strength of the relationship between children's vocabulary and understanding of cardinality.

Keywords

Data Classification, Vocabulary, Cardinality Understanding, Knower-level, Proportion-correct

Introduction

As humans develop and grow, they acquire several skills that become important to their ability to function in daily life. Whether it be how to walk, or how to read and write, several essential skills developed at a young age are utilized for the rest of their lives. Two essential abilities humans acquire are the ability to communicate through language, and the ability to understand and perform basic numeric operations. The acquisition of these abilities generally happens during the infant and toddler stages of development, and is crucial for participation and success in formal education where these basic skills are built upon (Baroody et al., 2006; Pace et al., 2019).

The development of children's understanding of counting, quantity-based comparisons, and the meaning of number words is often referred to as the development of their number knowledge. Research has provided evidence for two core systems for representing number: the parallel individuation system, which allows individuals to create exact representations of quantities smaller than four, and the approximate number system, which allows individuals to create approximate representations for quantities greater than or equal to four (Feigenson et al., 2004). These abilities have been observed in infants and non-human animals, providing evidence that language is not necessary to use these core systems (Spelke, 2011). While these core systems on their own do not require language, it has been hypothesized that language may serve as a medium to combine these two core systems, allowing for exact representations of larger sets--what is referred to in the literature as an understanding of "cardinality" (Spelke, 2011; Carey, 2009). To give an idea of how this combination occurs, children learn the ordered count list in their language, and then learn how to link a subset of number symbols to their cardinal values. Children progressively learn the meanings of number words, starting with "one", "two", "three", and so on (Wynn, 1990). At some point after this (around three and a half to four years of age), children's understanding of how to count allows them to create exact representations of sets of all sizes, and express the numerosity of large sets, demonstrating an understanding of cardinality. Since counting and the use of cardinality are essential for performing simple and advanced operations on sets and Arabic numerals in a formal math education setting (Baroody et al., 2006), it is important for researchers to look at how language might support the development of cardinality understanding.

Grammatical number marking and parental language input affect cardinality understanding.

A few studies have provided evidence that learning the cardinal meanings of the early numbers in the count list (specifically "one", "two", and "three") is influenced by the grammatical structure of singular, dual, and plural markings in the language a child is learning. Sarnecka et al. (2007) compared children learning English and Russian, both of which have obligatory singular and plural grammatical markings when discussing one item versus many items, to children learning Japanese, which does not grammatically mark singular and plural. They found that Russian- and English-speaking children were better able to distinguish between one object and multiple objects compared to Japanese-speaking children of similar ages, and that this difference was not related to children's counting ability. Furthermore, Marusic et al. (2016) tested the cardinality understanding of children learning Slovenian, which, in addition to singular and plural, contains obligatory dual markings when talking about sets of exactly two items in certain dialects, but does not contain obligatory dual markings in other dialects. The children learning the dialects that included the dual markings learned the meaning of "two" more quickly than the children learning the dialects lacking dual markings; children who knew the meaning of two were also less frequent in the sample of children learning the dialects that did not have obligatory dual markers (Marusic et al., 2016). Sarnecka (2014) summarized the results of several studies looking at the relationship between grammatical marking of number in a child's language and their understanding of cardinality, and concluded that grammatical marking of number influences children's learning of the meanings of number words, providing strong evidence for a relationship between the morphosyntax a child is exposed to during language acquisition and the timescale in which their cardinality understanding develops.

In addition to grammatical number marking influencing children's development of cardinality understanding, Gunderson et al. (2011) analyzed parents' number talk of children less than 2 years and 6 months of age. In their analysis, number talk referred to any use of the words "one" through "ten" in a numerical context, which meant counting with or without objects present, and expressing cardinal values of numbers with or without objects present; set sizes varied categorically between small sets (quantities 1 to 3) and large sets (quantities 4 to 10). When the children reached 3 years and 10 months of age, their understanding of cardinality was measured and analyzed alongside their parents' number talk. Their findings showed that parents' number talk regarding the counting and cardinal labeling of sets of present objects was correlated with children's later understanding of cardinality; this was especially true for the large set sizes. Gibson et al. (2020) further explored this relationship, and found parent number talk to be causally related to children's understanding of cardinality.

Based on these previous studies, it appears that language structure, children's language input, and children's language development are all related to their development of cardinality understanding. The questions that continue to be investigated are: how strong is this relationship, what is it contingent upon, when in development is language important for developing the understanding of cardinality, and what aspects of language are important. One aspect of language that researchers have extensively studied in regards to this relationship is children's general[1] vocabulary. There is a discrepancy among these studies as to whether or not a relationship between children's vocabulary and their understanding of cardinality exists; some studies provide data analysis that supports the existence of this relationship (Negen & Sarnecka, 2012), while other studies provide data analysis opposing this conclusion (Ansari et al., 2003; Hornburg et al., 2018).

Conflicting findings on the relationship between vocabulary and cardinality understanding.

When examining the reason for why studies have come to different conclusions regarding the existence of a relationship between vocabulary and children's understanding of cardinality, a main factor that must be considered is the various ways in which the data from cardinality tasks are classified. One task used in several past studies to evaluate children's understanding of cardinality is the Give-a-Number task ("Give-N"), where children are asked to form a subset of a specific quantity of objects from a larger set. Generally, there are two main ways in which Give-N data are classified: the knower-level classification and the proportion-correct classification. The knower-level classification, tracking the highest quantity reliably given by the participant, is discrete and ordinal. In contrast, the proportion-correct classification, which tracks the number of trials where the quantity was correctly given out of the total number of trials administered, is continuous. These classifications require the use of different analytical methods; this, as well as the different distributions of the data resulting from the two classifications, can have a profound effect on conclusions. *This work explores how these different data classifications (knower-level versus proportion-correct) may lead to contrasting analytical conclusions about the relationship between children's vocabulary and understanding of cardinality. In short, we found the classification of the Give-N data to have no effect on the existence of this relationship in our own dataset; the receptive vocabulary scores were significantly correlated with the knower-level and proportion-correct Give-N scores.*

[1] General vocabulary refers to both expressive and receptive vocabulary. Some studies look at one or the other, and some look at both.

Negen and Sarnecka (2012) conducted a similar exploration: their study tested children on Give-N, and related their Give-N scores to measures of both receptive vocabulary, as measured by the Peabody Picture Vocabulary Test (PPVT-3; Dunn & Dunn, 1997), and expressive vocabulary, as measured by the Woodcock-Johnson Picture Vocabulary Test from the Woodcock-Johnson II-R (Woodcock & Johnson, 1985). Partialling out age, they found significant correlations between the Give-N scores and vocabulary scores, both receptive and expressive, respectively, of their participants, thus providing evidence that there is a relationship between vocabulary scores and understanding of cardinality. This contrasted with the results of Ansari et al. (2003), which did not find a relationship between receptive vocabulary, measured using the British Picture Vocabulary Scales (BPVS; Dunn et al., 1982), and understanding of cardinality, reflected by a lack of a significant correlation between their participants' receptive vocabulary scores and Give-N scores. One of the main ways in which these studies differed was in how they classified their data; Negen and Sarnecka (2012) used the knower-level classification, while Ansari et al. (2003) used the total number of objects correctly given, which is similar to a proportion-correct classification. As an extension of their study, Negen and Sarnecka (2012) generated a large set of simulated participant data, including simulated ages, receptive vocabulary scores, and Give-N scores. They analyzed the Give-N data using both types of classifications, and ran correlational analyses with the generated vocabulary scores (partialling out age). These analyses were repeated 10,000 times for each classification type, and they recorded whether the correlations were significant between the receptive vocabulary and Give-N scores. They found that the knower-level classification resulted in a significant correlation 12% more often than the proportion-correct classification, indicating that studies examining the relationship between receptive vocabulary and cardinality understanding using the knower-level classification are more likely to detect a relationship. It is important to note that when using the proportion-correct classification they did find a significant correlation in some of their simulations, just not as often as when they used the knower-level classification. Ultimately, they concluded that the knower-level classification should be used to classify Give-N scores as this classification maximizes the statistical power of analyses when using Give-N as a measure of cardinality understanding, and is thus more likely to correlate significantly with receptive vocabulary measures.

In addition to the statistical evidence provided by Negen and Sarnecka (2012) that the knower-level classification provides a more powerful measure of children's understanding of cardinality, it is important to theoretically consider how the knower-level classification of Give-N data showcases a child's number knowledge in comparison to the proportion-correct classification. The knower-level classification has a more rigorous criterion for determining children's knowledge of number word meanings as it ensures children only receive credit for quantities they reliably get correct during testing; children are always given multiple trials on the quantities they are assessed on, and need to provide the correct quantity on a majority of the trials[2]. This stricter criterion limits the possibility of children giving correct quantities by chance, and receiving credit for knowing a number they may not actually know. The proportion-correct classification, however, is influenced by trials where children get quantities correct by guessing, or by chance. It also collapses the correct responses over the total number of trials, resulting in only seeing what proportion of responses they got correct, and this may be obscuring which quantities children actually appear to know or not know.

Theoretically, the points made above support the idea that the knower-level classification of Give-N data may be a more precise way of measuring children's cardinality understanding, and the statistical evidence from Negen and Sarnecka (2012) points towards the knower-level classification as more powerful for detecting significant correlations with other variables. With that being said, we cannot definitively say that the knower-level classification is more accurate than the proportion-correct classification for capturing children's understanding of cardinality. What may be more appropriate to say is that these two classifications showcase children's cardinality understanding differently. To give a few specific examples of how these two classifications showcase a child's cardinality understanding differently, we can consider a Give-N task with 18 trials: 3 trials each for quantities 1 through 6. For knower-level scoring purposes, a child must get two out of the three trials for a quantity correct to receive credit for knowing that number, and their knower-level will be determined by the highest number in which

[2] Past literature and our study require two correct trials out of the three given for each quantity to get credit for knowing the meaning of a number.

they get credit for knowing while also getting credit for all the numbers below. For proportion-correct scoring purposes, the child's total number of correctly given responses will be divided by 18, the total number of trials administered. In a given scenario, the estimation of a child's cardinality understanding based on their Give-N success may change depending on what classification is used. **Table 1** provides some scenarios to demonstrate this idea.

Performance scenario	Knower-level	Proportion correct	Estimation of cardinality understanding
A child gets 2 out of 3 of the trials correct for each quantity	6-knower	0.67	The child performs at ceiling using the knower-level classification, but only at 0.67 using the proportion-correct classification. The proportion-correct classification may be underestimating the child's cardinality understanding.
A child gets 2 out of 3 of the trials correct for quantities 1 through 3, and 1 out of 3 of the trials correct for quantities 4 through 6	3-knower	0.50	The child's performance is average according to both classification types, and their cardinality understanding is estimated similarly.
A child gets 1 out of 3 of the trials correct for each quantity	Pre-knower[3]	0.33	The child performs at floor using the knower-level classification, but at 0.33 using the proportion-correct classification. The proportion-correct classification may be overestimating the child's cardinality understanding.

Table 1: Examples of different performance scenarios in a Give-N task and corresponding estimations of cardinality understanding according to knower-level and proportion-correct classifications.

The first and third example scenarios in Table 1 show that the classification type affects where the child falls in the distribution of Give-N performance for a set of data, and thus how their understanding of the meaning of number words is viewed. The second example in Table 1

[3] Pre-knowers do not meet the task's criteria for knowing the meaning of any number words, including "one".

shows that both classification types can showcase a child's cardinality understanding the same. The fact that these two characterizations sometimes align and other times differ makes it difficult to determine which one is accurately capturing children's understanding of cardinality through Give-N.

Despite the fact that these two classifications capture children's understanding of cardinality differently, and Negen and Sarnecka's (2012) assertion that the knower-level classification is more statistically powerful, the proportion-correct classification is still used to classify Give-N data. Hornburg et al. (2018) tested children on a battery of language and number tasks to see if their language ability facilitated their understanding of mathematical concepts. Their tasks included a modified Give-N task[4], and the expressive vocabulary subtest of the Clinical Evaluation of Language Fundamentals-Preschool-Second Edition (CELF:P2; Wiig, Secord, & Semel, 2004). They characterized their Give-N data using the proportion-correct classification, and did not find a significant relationship between these two measures, and thus concluded that there is no relationship between expressive vocabulary and cardinality understanding. Negen and Sarnecka (2012) measured the expressive vocabulary of their participants using the Woodcock-Johnson Picture Vocabulary Test. They did find a significant relationship between their participants' expressive vocabulary scores and their knower-level Give-N scores, and thus concluded that there is a relationship between expressive vocabulary and cardinality understanding. Given that Hornburg et al. (2018) and Negen and Sarnecka (2012) used different measures of expressive vocabulary, different Give-N tasks, and different classifications of their Give-N data, it is difficult to reconcile why they came to different conclusions regarding the existence of a relationship; the difference in data classification may be one of the contributors to this discrepancy.

Current study's research aim, significance, and hypotheses.

This study aims to further explore how different ways of classifying data can influence outcomes in order to help explain the reasons for discrepancies in past studies on the relationship between vocabulary and cardinality understanding. The two main conclusions from previous

[4] Their "Give-N" task tested children's knowledge of the number words 3, 4, 6, 8, 11, and 16 one time each.

research regarding the investigated relationship are completely contradictory; it is important to determine which conclusion is correct, and what factors the conclusion depends on. A better understanding in this domain will help to inform parents and educators on what type of interventions may be needed for children struggling with their development of general vocabulary or cardinality understanding. The importance of investigating the reasons for these discrepancies can also be applied on a more global scale. In any type of scientific study, the way that data is classified and analyzed can change the types of conclusions made about studies. The results of scientific studies drive many clinical, political, and social decisions and policies, so therefore, ensuring that the data collection and analysis process is as objective and accurate as possible is critical for making the best decisions for members of all national and global communities.

This work utilized Give-N data collected as part of the Study of Language and Math (SLaM), funded by an NSF CAREER award to Dr. Marie Coppola at the University of Connecticut. The goal of this project is to see if the classification of the Give-N data affects the significance of the correlation with a receptive vocabulary measure, and ultimately the conclusion regarding the existence of a relationship between vocabulary and cardinality understanding. We measured participants' receptive vocabulary scores using the Peabody Picture Vocabulary Test (PPVT-4; Dunn & Dunn, 2007), and separately correlated those receptive vocabulary scores with both the knower-level and proportion-correct classifications of participants' Give-N scores. Based on Negen and Sarnecka's (2012) analyses, we expect to find a significant correlation between PPVT and Give-N scores when the Give-N data are characterized using the knower-level classification, and not find a significant correlation between PPVT and Give-N scores when the Give-N data are characterized as proportion-correct[5]. Our data will allow us to apply Negen and Sarnecka's (2012) comparison methodology using real data (as opposed to their simulated data) and determine whether data classification affects the correlation between children's PPVT and Give-N performances, and the overall relationship between receptive vocabulary and understanding of cardinality as measured by these tasks.

[5] Negen & Sarnecka found that PPVT scores were significantly correlated with Give-N proportion-correct scores in some of their simulated data sets. Therefore, it is possible that we may find a significant correlation when using the Give-N proportion-correct classification.

Additionally, we discuss other reasons why past research studies have come to different conclusions regarding the relationship.

Methods

The SLaM project tested deaf/hard-of-hearing and typically hearing children to examine the relationship between language and number development in children who have variable language experience, due in part to their hearing levels and consequent ability to access spoken language, and whether and when they were exposed to a sign language. For the current thesis project, we only use the typically hearing participants' data for analyses. Children completed several tasks, including the PPVT to provide a measure of their receptive vocabulary, and the Give-N task to test their understanding of cardinality.

Participants.

The participants for this study were 47 children (31 female), with an average age of 4;9 years (*SD* = 9.6 months, range = 3;0 to 6;6 years). All the children spoke English at home, had no known disabilities, and were recruited from schools across the state of Connecticut.

Procedure.

PPVT Task

We adapted Form B of the PPVT-fourth edition (PPVT-4; Dunn & Dunn, 2007) to obtain receptive language measures for our participants. A 13-inch Macbook Pro laptop was placed next to the test booklet in front of the child, and pre-recorded videos of an experimenter saying the target word (in English) were played on the laptop for each trial. The task was video recorded for archival and scoring purposes after the testing was completed[6].

[6] Our administration of the PPVT deviates from the conventional way of administration. These data were collected as part of the SLaM study which tested deaf/hard-of-hearing and typically hearing participants. In order to adapt the PPVT for the signing deaf/hard-of-hearing participants, the task was administered using a laptop with videos of the vocabulary words being signed. This methodology was also used for the hearing participants (and the deah/hard-of-hearing participants who were acquiring spoken English), showing them videos of an experimenter speaking the words, in order to make the task administration comparable across groups.

During the task, children were presented with a page containing four pictures of various objects, or people performing actions. The child was asked to point to the picture in the book that most accurately matched the given target word. If the child did not hear the target word, or did not seem to understand the word, the video of the experimenter saying the target word was replayed. After replaying the video, if the child did not respond within 10 seconds, the experimenter prompted them to do their best and choose the picture they think best matched the word. If a child seemed to be progressing too quickly, or aimlessly pointing at the pictures, the experimenter urged them to consider all four options before answering. As the test progressed, the target words became more difficult.

The target words were broken down into sets, each containing 12 target words. The words were numbered from 1 (the first target word in the first set) to 168 (the last target word in the final set). The criteria used to determine a child's starting set corresponded with their age. If a child produced more than one error in their starting set, the experimenter tested them on the previous set. This was done until a basal set was established for the child (i.e. a set where the child had at most one error). After completing the basal set, the experimenter progressed through each set until the child produced 8 errors in a set, and the experimenter ended the task. The maximum number of sets a child could complete was 14 (starting at set 1 and ending after completing set 14).

The starting set, basal set, total number of target words completed, ceiling set, and total number of errors were recorded for each child. Each child's PPVT raw score was calculated by subtracting the child's total number of errors from the number corresponding to the highest item completed in their ceiling set.

Give-a-Number ("Give-N") Task

Give-N was used to measure children's understanding of cardinality. Each child was tested on three trials each of quantities 1 through 6, presented in a pseudorandomized, predetermined order (shown below in **Table 2**). In this task, children were presented with a bowl and 20 multicolored plastic fish. The experimenter asked the participant to put *X* (the target quantity) fish in the bowl. Once the child had put fish into the bowl, they were asked "Is that *X*?"

The experimenter then dumped the fish onto the table and lined them up. The children were asked "Can you count to make sure there are *X*?" If the child realized they made a mistake while counting, they were given an opportunity to fix their mistake; however, regardless of children's success at counting, if the child said the set produced was the target number, the experimenter moved on to the next trial.

The task was scored by undergraduate research assistants. For each trial, a binary code ('1' for correct and '0' for incorrect) was used to determine if a child produced the quantity asked for. Children received credit for knowing a given quantity if they produced a set of the correct quantity on at least two of the three trials for that quantity.

Trial Number	Quantity assessed	Trial Number	Quantity assessed	Trial Number	Quantity assessed
1	1	7	3	13	1
2	4	8	2	14	5
3	2	9	1	15	4
4	3	10	5	16	3
5	6	11	4	17	2
6	5	12	6	18	6

Table 2: The order of Give-N trials administered to participants.

Give-N Classifications and Analyses.

The primary question we will address is: Does the way we classify Give-N data affect the relationship between receptive vocabulary scores and Give-N scores? To ensure this question could be answered, our Give-N data were classified in two ways: knower-level and proportion-correct. A child's knower-level corresponded to the highest quantity in which they received credit for while also receiving credit for all the quantities below (credit being getting two out of the three trials for a quantity correct), and this acted as an ordinal measure of each child's performance on Give-N (this is similar to the knower-level classification used by Negen & Sarnecka, 2012). The proportion-correct classification was calculated by dividing the number of trials correctly produced by the total number of trials presented to the child, and this acted as a continuous measure of each child's Give-N performance (this is analogous to the proportion-correct calculation used by Hornburg et al., 2018). The PPVT scores were kept in the form of raw scores, following Negen and Sarnecka (2012).

To examine the relationship between children's receptive vocabulary and Give-N performance using both classifications of data, the PPVT scores and Give-N scores under both classifications were correlated. The correlations were compared with the goal of providing an answer to the question of whether data classification affects the relationship between receptive vocabulary and Give-N scores.

Results and Discussion

Did a relationship between receptive vocabulary scores and both classifications of Give-N scores exist in our data?

This thesis examined the relationship between receptive vocabulary and understanding of cardinality by seeing if the correlation between participants' PPVT scores and their Give-N data was affected by the classification of the Give-N data. **Figure 1** shows the Give-N data graphed against the age of the participants for each Give-N classification type.

For both the knower-level classification, and the proportion-correct classification, the Give-N scores generally increased with age until around 4 and a half or 5 years where there appeared to be a ceiling effect. The ceiling effect observed in these graphs is expected given that previous research has found that around 4 years of age, children are able to generalize their understanding of the meaning of number words over the entire count list (Wynn, 1992). The bulk of the participants in our data set (81%) are 4 years of age or older, and therefore were likely to perform at ceiling in the Give-N task. The ceiling effect may be more pronounced in the proportion-correct Give-N data since there is a clearer transition from a line with positive slope to a line with no slope (right) compared to the knower-level data (left).

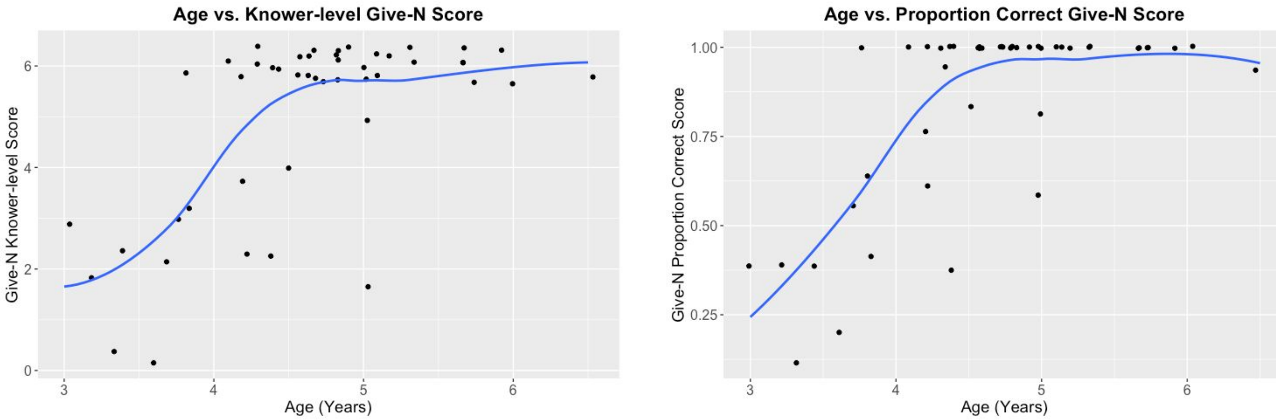

Figure 1: Scatterplots showing participant ages graphed against their respective Give-N knower-level scores (left) and Give-N proportion-correct scores (right). The graphs are fit with loess curves (in blue). The points in the left panel appear to be slightly above or below knower-level values because jitter was used to generate these plots.

Figure 2 shows the participants' ages graphed against their PPVT raw scores. This graph did not show a ceiling effect, but rather a trend of PPVT raw scores increasing continuously with age. The end of the trend line starts to curve downward due to the fact that there are few data points for participants over the age of 5.5 years. More data points for participants in this age range may reveal a continuation of the upward trend.

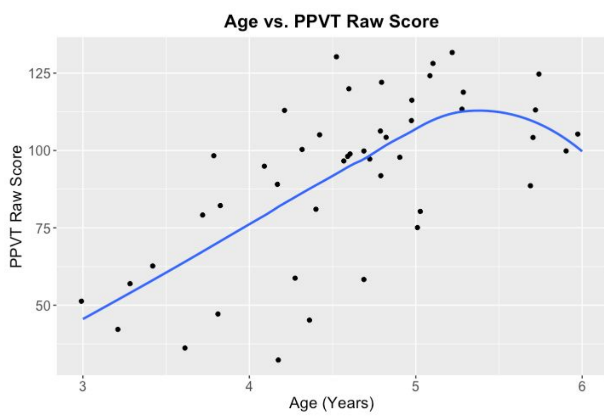

Figure 2: Scatterplot showing participant ages graphed against their respective PPVT raw scores. The graph is fit with a loess curve (in blue).

When comparing how the Give-N scores from each classification type individually correlated with the PPVT raw scores, the first step was to see if the two classifications themselves were correlated. If these two types of classifications don't correlate with each other, then there would already be strong evidence pointing towards classification type being a contributing factor to different findings regarding the relationship between vocabulary and Give-N scores. The correlation (partialling out age) between the two classifications of our participants' Give-N data was strong ($\rho=0.919$, $p<0.001$). In the introduction, it was mentioned that classifying the data using the knower-level and proportion-correct classifications may give a different idea of where a child lies on the distribution of Give-N scores, but the strong correlation between the classification types tells us that each captures children's knowledge of the meaning of number words equivalently, at least for participants in our dataset (but see discussion of how age impacts Give-N distribution in the "Exploring the discrepancies" section below).

Figure 3 shows the participants' PPVT raw scores graphed against their Give-N scores of both classifications. In the graphs for each classification, there appears to be a ceiling effect that begins around 100 on the x-axis of PPVT raw scores. The ceiling effect shown is likely due to the fact that many older children performed at ceiling on Give-N (as shown in Figure 1), yet PPVT scores continuously increased with age (as shown in Figure 2), so it is not expected that there be a continuously increasing relationship between PPVT and Give-N scores across our sample.

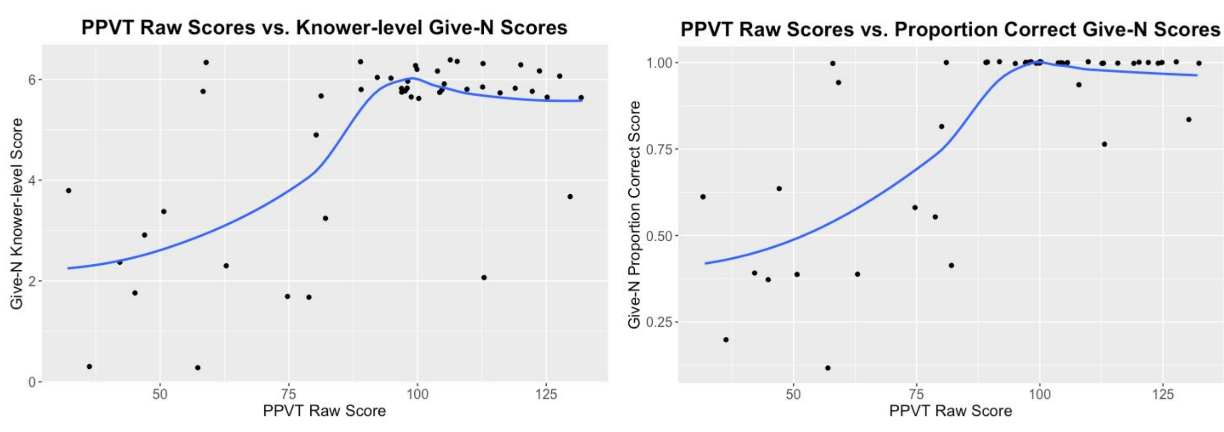

Figure 3: Scatterplots showing participants' PPVT raw scores graphed against their respective Give-N knower-level scores (left) and Give-N proportion-correct scores (right). The graphs are fit with loess curves (in blue).

To determine whether participants' PPVT raw scores and their Give-N knower-level scores were significantly correlated, we ran a Spearman correlation[7] between the two variables[8]. There was a significant correlation between PPVT raw scores and Give N knower-level (ρ=0.313, p<0.03). Based on the knower-level classification we found that a relationship exists between the PPVT raw scores and Give-N scores. To determine whether participants' PPVT raw scores and their Give-N proportion-correct scores were significantly correlated, we ran a Pearson correlation between the two variables. There was a significant correlation between PPVT raw scores and Give-N proportion-correct (r=0.547, p<0.001). Based on the proportion-correct classification, we also found that a relationship exists between PPVT raw scores and Give-N scores.

The correlation between the PPVT raw scores and the proportion-correct classification is stronger than the correlation between the PPVT raw scores and the knower-level classification. In Table 1 in the introduction, we demonstrated how the classification types may capture children's cardinality understanding through Give-N differently, and example 1 in the table is relevant for discussion here. In that example, a child who performed at ceiling based on the knower-level classification only performed at 0.67 correct using the proportion-correct classification. In our study, given that the majority of our sample consisted of children over the age of 4 who are further along in their development of cardinality understanding , many of our children performed at ceiling on the Give-N task according to both classification types. However, we did have a few 6-knowers whose proportion-correct values ranged between 0.67 and 1.00. Because of this, the proportion-correct classification may be giving a more nuanced look at children's Give-N performance, and thus may be providing a better insight into the relationship between receptive vocabulary and Give-N scores in our study.

Overall, we found no difference in the existence of a relationship between receptive vocabulary and Give-N scores based on the classification of our Give-N data. We did find that

[7] We interpreted our Give-N knower-level data as ordinal, which called for the use of a Spearman correlation. Negen and Sarnecka (2012) used a Pearson correlation with their Give-N knower-level data, but they do not state why within their text. See pages 25-26 for further discussion of how analyses chosen may affect conclusions.
[8] In order to reduce the effects of participant ages on the correlation, and to be consistent with the analytical methodologies of previous research studies (Ansari et al., 2003, Negen & Sarnecka, 2012, and Hornburg et al., 2018), we partialed out age in all correlations between PPVT and Give-N scores.

the correlation was stronger between PPVT raw scores and the proportion-correct classification. Although both were significant, classification type is still an important factor to research regarding its effect on the relationship between receptive vocabulary and Give-N scores. This is due to the fact that one may have more of an effect than the other, or they may have different strengths of effect/no effect at all based on other factors in the study, including a child's age, for instance. Also, discrepancies still exist within the research regarding this relationship, including the results of this study.

How does this compare with previous research?

The motivation behind our research was not only to test the relationship between receptive vocabulary and Give-N scores using data we collected, but also to discuss reasons why past studies that have tested the relationship between vocabulary scores and Give-N scores have come to different conclusions about the existence of a relationship. Thus, it is important to compare the results of our study to those of other studies.

Beginning with Negen and Sarnecka (2012), children were administered the PPVT and a Give-N task that tested their knowledge of quantities of 1 through 6. They found a relationship between receptive vocabulary and Give-N knower-level scores (r=0.546, p<0.01). This result is consistent with ours, as we found a significant correlation between our PPVT receptive vocabulary scores and Give-N knower-level scores (ρ=0.313, p<0.03). Their correlation was stronger than ours, signifying a stronger relationship in their experiment; this may be due to the fact that their participants were earlier in their development of cardinality understanding compared to our participants.

Ansari et al. (2003) tested participants' receptive vocabulary using the British Picture Vocabulary Scales (BPVS; Dunn et al., 1982), and their understanding of cardinality using a Give-N task with quantities 1 through 6. They did not find a significant correlation between their participants' receptive vocabulary scores and proportion-correct Give-N scores (r=0.049). This result is inconsistent with our data, as we found a significant correlation between our participants' receptive vocabulary scores and Give-N scores using the proportion-correct classification (r=0.547, p<0.001).

Hornburg et al. (2018) tested participants' expressive vocabulary using the CELF:P2, and used a modified Give-N task to test children's understanding of cardinality (see footnote on page 9 for a description of their Give-N task). Their analyses indicated a significant correlation between their participants' expressive vocabulary scores and their proportion-correct Give-N scores (r=0.44, p<0.05), however, when partialling out the effects of other variables, including age, sex, parent education, and children's mathematical language (e.g., their understanding of quantitative and spatial words) using a mixed effects regression, expressive vocabulary was no longer significantly related to Give-N scores. This result may be inconsistent with our findings, which indicated a relationship between receptive vocabulary and Give-N performance.

Exploring the Discrepancies.

The fact that some studies find a relationship between their vocabulary measure and Give-N data and some do not is consistent with Negen and Sarnecka's (2012) findings that the way Give-N data are classified may affect this relationship. In their experiment, they simulated 10,000 data sets that included age, receptive vocabulary, and Give-N scores of both classifications (refer back to page 6 for experiment details), and found a significant correlation between PPVT and Give-N scores 12% more often using the knower-level classification compared to the proportion-correct classification. They concluded that the knower-level classification is a more powerful characterization of Give-N data, and that it was more likely to result in a significant correlation with vocabulary scores. Our study found a significant correlation between PPVT and Give-N knower-level scores, supporting the argument that the knower-level classification is powerful for detecting relationships between Give-N scores and other variables, specifically receptive vocabulary.

Ansari et al. (2003) used the proportion-correct classification for their Give-N task, and did not find a relationship between receptive vocabulary and their participants' Give-N scores. Similarly, Hornburg et al. (2018) used the proportion-correct classification for their modified Give-N task, and did not find a relationship between their participants' expressive vocabulary and Give-N scores. Using the idea put forth by Negen and Sarnecka (2012) about the knower-level classification being more powerful, if we could reclassify the Give-N data from

Ansari et al. (2003) and Hornburg et al. (2018) to the knower-level classification, we might have found a relationship to exist.

The discrepancies between these studies may also be due to sampling limitations. Our study found a significant correlation between PPVT and Give-N proportion-correct scores, contrary to both Ansari et al.'s (2003) findings, and Hornburg et al.'s (2018) findings. Negen and Sarnecka (2012) may have found significant correlations between PPVT and Give-N scores more often using the knower-level classification as mentioned above, but they did find significant correlations using the proportion-correct classification in some subset of their simulated data sets. Thus, the fact that we found a correlation between vocabulary scores and proportion-correct Give-N scores, and Ansari et al. (2003) and Hornburg et al. (2018) did not, may serve as a real example of what Negen and Sarnecka (2012) found in their simulated data: the proportion-correct classification is sometimes significantly correlated with receptive vocabulary scores, but not always. Through repeated experimentation with their participants and procedures, Ansari et al. (2003) and Hornburg et al. (2018) may have found significant correlations between their participants' vocabulary scores and Give-N proportion-correct scores.

In addition to data classification and sampling explaining the discrepancies between these studies' findings, there are several other differences with respect to how each study collected their data that may also have contributed to the different conclusions these studies reached regarding the existence of a relationship between vocabulary and cardinality understanding.

First, there are several different ways to administer the Give-N task. Our study, and Ansari et al. (2003), tested children three times each on quantities 1 through 6, for a total of 18 trials. Although our trials were presented in a predetermined, pseudorandom order, and Ansari et al.'s (2003) were presented "randomly[9]", the participants in our studies were given the same number of trials for the same quantities. Negen and Sarnecka (2012) used a titrated method to test children's understanding of cardinality, meaning the number of trials and the quantities a child was tested on depended on their success in previous trials. Once a child got two trials correct for a given quantity, and two trials incorrect for the subsequent quantity in the count list,

[9] Ansari et al. (2003) do not make it clear if every child received a uniquely randomized order of Give-N quantities, or if they used a similar method as us and randomized a single order of trials beforehand that was administered to each child.

they were placed in a knower-level category that reflected their highest correct quantity. While the order of the trials was predetermined, and the quantities tested ranged from 1 to 6, the total number of trials differed between children depending on their success in completing the task. Hornburg et al. (2018) used a modified Give-N task and tested children's knowledge of quantities 3, 4, 6, 8, 11, and 16. Each quantity was administered once for a total of 6 trials. The difference in the actual quantities tested, in the number of trials a child was administered, both for specific quantities and in total, and whether or not a titrated method was used in testing may change the way children's understanding of cardinality is measured. Certain ways of administering the Give-N task may give different pictures of children's understanding of cardinality, and from a big picture standpoint may result in varying conclusions regarding the existence of a relationship between vocabulary and cardinality understanding.

Second, the studies examined in this section used different tests to assess children's vocabulary knowledge. Negen and Sarnecka's (2012) study used the Woodcock-Johnson Picture Vocabulary Test to gather expressive vocabulary scores for their participants, and the PPVT to gather receptive vocabulary scores for their participants. Our study also used the PPVT to gather receptive vocabulary measures for our participants. Ansari et al. (2003) utilized the BPVS as a receptive vocabulary measure for their participants, and Hornburg et al. (2018) used the CELF:P2 as an expressive vocabulary measure in their study. The type of vocabulary measure (expressive vs. receptive), and the specific assessments used, may have an effect on whether or not a relationship is observed between vocabulary and cardinality understanding. Negen and Sarnecka (2012) examined the relationship between their vocabulary measures (expressive and receptive) and Give-N scores for each type of vocabulary measure, and they found a relationship to exist in both instances. This suggests that we do not necessarily expect to see a difference regarding the relationship between expressive/receptive vocabulary and Give-N, but it is something that needs to be considered and studied in a more comprehensive and systematic way.

Lastly, the age of the participants in a study may have an effect on the conclusions gathered by researchers regarding this relationship, and this is something we will explore further in this section. At around 3 and a half to 4 years of age, children demonstrate an understanding of cardinality, and can reliably and exactly represent and produce sets of large quantities (Wynn

1990). Thus, age can have an effect on a child's Give-N performance, especially in testing scenarios where the largest quantity tested is 6.

Participant age range affects the distribution of Give-N performance, and may in turn affect whether or not a relationship is found between vocabulary and Give-N scores. Looking at **Table 3**, and comparing the knower-level breakdown from the Negen and Sarnecka (2012) study (shaded in dark blue) to the knower-level breakdown of our study (shaded in red), we see that only 42% of their participants were classified as 5- and 6-knowers (grouped into one group called CP-knowers) while 72% of our participants were classified as 5- or 6-knowers. Negen and Sarnecka (2012) included participants in a younger age range than the age range of participants in our study, and their data included fewer children who performed at ceiling on the Give-N task. Additionally, although both of our studies concluded that a relationship exists between PPVT scores and Give-N knower-level scores, the relationship based on their correlation between their participants' PPVT scores and Give-N knower-level scores was stronger than ours. The younger age range of their participants, and the smaller proportion of participants who performed at ceiling on Give-N, may have resulted in more variability in Give-N scores, which would translate into a stronger relationship between Give-N scores and similarly varied vocabulary scores.

The idea that age range affects the distribution of Give-N performance is also reflected specifically within our data. When our participants were subset into age ranges that matched those of the Negen and Sarnecka (2012) study, the Ansari et al. (2003) study, and Hornburg et al. (2018) study (in light blue, green, and purple, respectively), we observed different proportions of children in the knower-level categories. For instance, when our data was subset to match the Negen and Sarnecka (2012) age range, only 54% of children performed at ceiling, whereas when our data was subset to match the Hornburg et al. (2018) age range, 76% of children performed at ceiling. Negen and Sarnecka (2012) included participants in a younger age range than the age range of participants in Hornburg et al. (2018), and there are many fewer children who performed at ceiling in the younger age range. The Ansari et al. (2003) age range subset of our data also includes younger children, and is more similar to the Negen and Sarnecka (2012) age range. Sixty-three percent of our participants in the Ansari et al. (2003) age range subset performed at ceiling, compared to 76% of our participants in the Hornburg et al. (2018) subset. These two comparisons using subsets of our study data demonstrate how the distribution of Give-N performance can be different based on the age range of the children being tested, which in turn may have an effect on whether a relationship is observed between Give-N scores and general vocabulary.

Knower Level:	**Pre**	**1**	**2**	**3**	**4**	**5**	**6**	**Total ***
Our Data, full age range (3.00 - 6.50 years): Number of participants	2	0	6	3	2	1	33	47
Our Data, full age range (3.00 - 6.50 years): Proportion of total participants	0.04	0	0.13	0.06	0.04	0.02	0.70	1.00
Negen & Sarnecka (2012) Data (2.50 - 5.00 Years): Number of participants	2	5	7	4	1	14 (called "CP - knowers")		33
Negen & Sarnecka (2012) Data (2.50 - 5.00 Years): Proportion of total participants	0.06	0.15	0.21	0.12	0.03	0.42		1.00
Our Data, Negen & Sarnecka (2012) age range^: Number of participants	2	0	5	3	2	0	14	26
Our Data, Negen & Sarnecka (2012) age range: Proportion of total participants	0.08	0	0.19	0.12	0.08	0	0.54	1.00
Our Data, Ansari et al. (2003) age range^ (2.50 - 5.25	2	0	6	3	2	1	24	38

Years): Number of participants								
Our Data, Ansari et al. (2003) age range (2.50 - 5.25 Years): Proportion of total participants	0.05	0	0.16	0.08	0.05	0.03	0.63	1.00
Our Data, Hornburg et al. (2018) age range (3.52 - 6.03 Years): Number of participants	1	0	4	2	2	1	32	42
Our Data, Hornburg et al. (2018) age range (3.52 - 6.03 Years): Proportion of total participants	0.02	0	0.10	0.05	0.05	0.02	0.76	1.00

Table 3: Knower-level frequency of the age group in our study, of the subsets of the age group in our study, and of the age group from the Negen and Sarnecka (2012) study.
*Proportion totals of some rows may not add up exactly to 1.00 due to rounding.
^Our dataset does not include any children younger than 3.00 years.

The figures presented at the start of the Results and Discussion section help to further demonstrate how age may contribute to the relationship between vocabulary and understanding of cardinality. Referring back to Figure 1, we noticed ceiling effects in the graphs of participant ages vs. their Give-N scores of both classification types. Due to the fact that older children can more reliably and exactly represent and produce sets of large quantities (Wynn 1990), they will often perform at ceiling in a Give-N task that only tests knowledge of the meaning quantities up to 6; the design of the task accurately captures older children's understanding of cardinality, which may be further along in development, and their ability to know the meaning of number words less than or equal to 6. Figure 2, a graph of participants' ages vs. their PPVT raw scores, showed a continuously increasing relationship between age and PPVT raw scores, which makes sense considering the PPVT is intended for a wide age range (PPVT-4 is designed for ages 2;6 years to 90;0 years according to Dunn & Dunn's (2007) PPVT-4 manual), and is specifically designed for children to not perform at ceiling; the starting set in the PPVT task is influenced by a participant's age, and the words become progressively more difficult as the task goes on, making it much more challenging to reach the ceiling item. This design is representative of how we learn vocabulary throughout our lives. Our vocabulary never reaches a "ceiling" as we are always learning new and more challenging words. These two ideas together lead to a ceiling effect in the graphs of PPVT raw scores vs. Give-N scores of each classification type in Figure 3. There is not a continuously increasing relationship between PPVT and Give-N scores across our sample. From this, we can conclude that it is possible that age may affect the relationship between vocabulary and cardinality understanding up to a certain point, specifically before children are able to reliably and exactly represent and produce sets of large quantities.

Not only can age alone affect the relationship between general vocabulary and Give-N performance, but there may also be an interaction between data classification and participant ages that affects the relationship. It is likely that studies which include children in older age ranges will have more children who perform at ceiling on a Give-N task that only tests up to 6. In the introduction, we mentioned how different data classifications can change where a specific child falls on the distribution of Give-N scores, relative to other children, depending on their task performance. To recall an example from this table, in a Give-N administration scenario similar to the one employed in our study, a child who gets 2 out of the 3 trials correct for each quantity would be classified as a 6-knower, but only have a 0.67 proportion-correct score. In this example, the child performs at ceiling under the knower-level classification, but only has a score of 0.67 using the proportion-correct classification, which is well under ceiling. The proportion-correct classification here may be underestimating the child's understanding of cardinality, as suggested in the introduction, or it may be giving a more nuanced look into their cardinality understanding. Children in older age ranges are more likely to have a better understanding of cardinality, which may improve their Give-N performance; if the knower-level classification is used with an older population of children, you may see a more pronounced ceiling effect as you may have many children classified as 6-knowers who got proportion-correct scores between 0.67 and 1.00. Thus, when working with populations of older children, the proportion-correct classification may be more appropriate to allow a range of variability in the data. Likewise, the knower-level classification may provide a better picture of younger children's Give-N data. As mentioned above, Negen and Sarnecka (2012) found using simulated data for a population of younger children that the knower-level classification is more likely than the proportion-correct classification to result in a relationship between receptive vocabulary and cardinality understanding.

Further considerations.

When comparing the conclusions from the studies above in relation to our study, there were several factors that came into play as potential reasons why the studies reached different conclusions. Differences in the Give-N task methodology, in terms of how the task was administered, what quantities were tested, and what criteria were used to score the participants, may have impacted the distribution of participants' Give-N scores. The type of data generated by the scoring process, and the necessary analyses may also have impacted whether or not a relationship was observed between the Give-N data and vocabulary measures. Lastly, the age range of the participants may have been a contributing factor to whether or not a relationship was found.

Finally, looking specifically at the knower-level classification, Give-N measured using knower-level seems like interval data, but children's progression through the knower-levels may not occur in an interval way. Treating Give-N data as interval data assumes an equal magnitude of difference between each knower-level (i.e. the difference in cardinality understanding between a 1- and 2-knower is the same as the difference in cardinality understanding between a 4- and 5-knower). However, when used as an outcome variable in ordinal logistic regression models, the knower-level classification violated the parallel lines assumption, meaning the effect of certain predictors was not equivalent for all transitions between knower-levels (E. Carrigan, personal communication, July 23, 2020). This provides some evidence that knower-level may not be a true interval measure, and suggests that we might find different relationships between vocabulary and cardinality understanding depending on where children are in their development. It is possible that language may be an important predictor for the transition between certain knower-levels, such as 1, 2, and 3, but not for the transition between other knower-levels, such as 4, 5, 6, and beyond.

Children's understanding of cardinality as demonstrated by Give-N knower-level may not be something that exists on a scale with equal spacing between knower-levels; there is not a way for us to quantify the difference in cardinality understanding between the knower-levels to ensure that they are equally scaled apart. Viewing Give-N knower-level data as ordinal may be more appropriate, as there is still order to the cardinality understanding between the knower-levels (i.e. a 4-knower understands cardinality more than a 2-knower) but the difference between the knower-levels is not necessarily evenly spaced apart. In our study, we viewed our Give-N knower-level data as ordinal, and thus we used a Spearman correlation to correlate the Give-N knower-level data with the PPVT data. Negen and Sarnecka (2012) may have assumed knower-level to be an interval data type since they used a Pearson correlation to correlate their Give-N knower-level data and PPVT data. The different type of correlation used may explain the difference in magnitude between our two correlations, which affects the perceived strength of the relationship between Give-N data and PPVT data in our two studies.

For future studies, it is important that researchers consider what type of measure knower-level is in order to carry out the appropriate analyses to get a more accurate understanding of the existence and strength of the relationship between general vocabulary measures and Give-N knower-level data.

Summary.

This study examined the relationship between children's understanding of cardinality as measured by Give-N and receptive vocabulary as measured by the PPVT. We looked at this relationship for two different classifications of the Give-N data: the knower-level classification, and the proportion-correct classification. We found there to be a significant relationship between both Give-N classification types and the receptive vocabulary data.

We also compared our findings to those of other studies in order to further explore the discrepancies across these studies. Several factors, including the way of administering the Give-N task, the type of language measure used and type of language being examined, the classification of the Give-N data, and the age range of the participants, may have contributed to the difference in conclusions between the studies. There may also be an interaction between age range and classification type that can affect the existence and strength of the relationship. Lastly, it is important to consider how different types of data require different analyses, which can also have an effect on a study's conclusion.

Future researchers looking at the relationship between children's general vocabulary and their understanding of cardinality as measured by Give-N should systematically alter the factors mentioned above in different experimental groups, and compare how they affect the outcome of the existence of a relationship to gain a better understanding of what factor(s) may be contributing to the discrepancy in findings between past studies. In instances where a relationship is found, future researchers should also look into how these factors affect the strength of the relationship, which is important for determining the practical importance and implications of such findings. Answering the questions posed here will help us have a better understanding of the relationship between general vocabulary and cardinality understanding, and will enable us to design effective intervention methods for children struggling with vocabulary or learning the meaning of number words.

The findings of this study, and research into the reasons for discrepancies between studies which have tested a similar relationship, have important implications for those who use the results of research, such as educators, parents, and clinicians. First off, replication of studies is extremely important in order to validate findings and ensure they are accurate. Scientific studies, especially those looking to replicate previous findings, need to consider all of the possible factors that can affect specific outcomes, and systematically alter these factors to experimentally determine their effects. Small changes between studies can lead to large differences in the outcomes between them. Along these same lines, one study or one finding should not be the basis for societal norms, policy changes, or curriculum design. We need to ensure the external validity of findings, or, in other words, that the findings of a study will hold true for the populations that resulting policies are applied to and not just for the research population. The process of establishing effective curriculum and policies is a difficult and slow process of extreme importance for the education of our children and operation of our society.

Acknowledgements

This work was supported by the National Science Foundation (grant number 1553589) and the University of Connecticut's Psychological Sciences Undergraduate Research Grant awarded to Bryne-Marie Sidney.

References

Ansari, D., Donlan, C., Thomas, M. S., Ewing, S. A., Peen, T., & Karmiloff-Smith, A. (2003). What makes counting count? Verbal and visuo-spatial contributions to typical and atypical number development. *Journal of Experimental Child Psychology*, *85*(1), 50-62.

Baroody, A. J., Lai, M., & Mix, K. S. (2006). Development of young children's early number and operation sense and its implications for early childhood education. In B. Spodek & O. N. Saracho (Eds.), *Handbook of research on the education of young children* (2nd ed., pp. 187–221). Mahwah, NJ: Erlbaum.

Carey, S. (2009). Oxford series in cognitive development: Vol. 3. *The origin of concepts*. New York, NY, US: Oxford University Press.

Dunn, L. M., Dunn, L. M., & Whetton, C. (1982). *The British Picture Vocabulary Scale: Manual for the short and long forms*. NFER-Nelson.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test–III*. New York: American Guidance service.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test: PPVT-4; Manual*. Pearson Clinical Assessment.

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody picture vocabulary test*. Pearson

Assessments.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in cognitive sciences*, *8*(7), 307-314.

Gibson, D. J., Gunderson, E. A., & Levine, S. C. (2020). Causal Effects of Parent Number Talk on Preschoolers' Number Knowledge. *Child Development*, 00(0), 1-16.

Gunderson, E. A., & Levine, S. C. (2011). Some types of parent number talk count more than others: relations between parents' input and children's cardinal‐number knowledge. *Developmental science*, 14(5), 1021-1032.

Hornburg, C. B., Schmitt, S. A., & Purpura, D. J. (2018). Relations between preschoolers' mathematical language understanding and specific numeracy skills. *Journal of experimental child psychology*, *176*, 84-100.

Marušič, F., Žaucer, R., Plesničar, V., Razboršek, T., Sullivan, J., & Barner, D. (2016). Does grammatical structure accelerate number word learning? Evidence from learners of dual and non-dual dialects of Slovenian. *PloS one*, *11*(8), e0159208.

Negen, J., & Sarnecka, B. W. (2012). Number‐concept acquisition and general vocabulary development. *Child development*, *83*(6), 2019-2027.

Pace, A., Alper, R., Burchinal, M. R., Golinkoff, R. M., & Hirsh-Pasek, K. (2019). Measuring success: Within and cross-domain predictors of academic and social trajectories in elementary school. *Early Childhood Research Quarterly*, *46*, 112-125.

Sarnecka, B. W., Kamenskaya, V. G., Yamana, Y., Ogura, T., & Yudovina, Y. B. (2007). From grammatical number to exact numbers: Early meanings of ‘one’,‘two’, and ‘three’in English, Russian, and Japanese. *Cognitive psychology*, *55*(2), 136-168.

Sarnecka, B. W. (2014). On the relation between grammatical number and cardinal numbers in development. *Frontiers in psychology*, *5*, 1132.

Spelke, E. S. (2011). Natural number and natural geometry. In *Space, time and number in the brain* (pp. 287-317). Academic Press.

Woodcock, R. & Johnson, M. B. (1985). *Woodcock-Johnson II Revised*. Itasca, IL: Riverside.

Wiig, E. H., Secord, W. A., & Semel, E. (2004). *CELF-Preschool-2: Clinical Evaluation of Language Fundamentals, Preschool*. Harcourt Assessment.

Wynn, K. (1990). Children's understanding of counting. *Cognition*, *36*(2), 155-193.

Wynn, K. (1992). Children's acquisition of the number words and the counting system. *Cognitive psychology*, *24*(2), 220-251.