# What Conceptions have US Grade 4-6 Students' Generalized for Formal and Informal Common Representations of Unknown Addends?

J. Matt Switzer
*Andrews Institute of Mathematics and Science Education*
*Texas Christian University*
j.switzer@tcu.edu

## Abstract

Research findings have established that students often struggle with mathematical symbols including common misconceptions for literal symbolic representations of variables but provide little evidence of when or how these misconceptions arise. This article reports findings from a study of grade 4-6 students' conception(s) for various representations of unknown addends [variables] commonly found in US elementary mathematics textbooks. Specifically, thirty-six US grade 4-6 students participated in two semi-structured task based interviews designed to explore their conception(s) of formal and informal representations of unknown addends as revealed by the types of numbers substituted across core mathematical tasks and task types. Findings showed that participants initially demonstrated a natural number bias and upon further questioning a whole number bias. Participants' number substitutions did not include common established difficulties for literal symbols exhibited by students in algebra courses and higher-level mathematics courses. In addition, even when providing the same types of number substitutions, participants attended to various attributes of the different representations for unknown addends.

*Keywords:*

Variable, representations, student thinking, mathematical symbolism, unknown addends

## Introduction

Successful completion of an *algebra* course, or equivalent, serves as a gatekeeper for students' future educational, professional, and economic opportunities (e.g., National Academy of Science, 2005; RAND Mathematics Study Panel & Ball, 2003). Recent high profile publications (e.g., National Academy of Science, 2005; National Mathematics Advisory Panel, 2008) have placed an increased emphasis on two aspects of teaching and learning algebra targeted at increasing student success. First, early mathematics education must *prepare* all students for future success in algebra. Second, experiences in algebra and higher-level mathematics courses must *support* all students in developing algebraic thinking and understanding, not just students who happen to be "successful" in current approaches to teaching mathematics.

Researchers have found that student difficulties with conventional mathematical symbols and representations can hinder students' preparation for and success in algebra. For example, students often view the equal sign as an operation (c.f., Alibali et al., 2013; Knuth, Stephens, McNeil, & Alibali, 2006; McNeil et al., 2006; Powell, 2012). Lobato, Ellis, and Muñoz (2003) described how four focusing phenomena ("goes up by" language, well ordered tables, graphing calculator, and uncoordinated sequences and differences) contributed to students unintended generalization of $y = mx + b$ as a difference. Booth (1984) found that students often encountered difficulties with the use of brackets to specify order of operations, demonstrated confusion between powers and products, and had an unwillingness to accept an algebraic expression as a solution. Lannin et al. (2013) found that first and second grade students who struggled learning mathematics exhibited difficulties representing problem situations with formal, informal and idiosyncratic written and physical representations.

Christou and Vosniadou (2012) found that students enrolled in algebra demonstrate a natural number bias in the numbers they substituted for literal symbols and that an initial understanding of number influenced students' knowledge acquisition processes in fundamental ways. Therefore, "it makes sense to assume that [initial understanding of number] would also influence the way students interpret the use of literal symbols" (p.5). In addition, Christou and Vosniadou's found that the "reference of literal symbols – the numbers literal symbols stand for is an important source of meaning, which influences performance in various mathematical tasks" (p. 4). In this paper, the author argues that extending Christou and Vosniadou's findings to include idiosyncratic and informal symbolic representations of variables, place holders, and unknown quantities commonly found in elementary school mathematics textbooks is equally important.

While research findings have identified common difficulties that students enrolled in algebra and higher-level mathematics courses have with literal symbolic representations of variables such as $y + y = 12$ and $a + b = 12$ (cf., Booth, 1984; Knuth, Alibali, McNeil, Weinberg, & Stephens, 2005; Macgregor & Stacey, 1997; Swafford & Langrall, 2000; Warren & Cooper, 2008), less is known about students' generalizations for idiosyncratic and commonly used informal representations of variables (e.g. __ + __ = 12, ☐+ Δ = 12, and ☐+ ☐= 12). Student generalizations for informal representations of variables, placeholders, and unknowns support their later generalizations and meaning for literal symbolic representations of variables and underpin early algebra (EA) and algebraic thinking, which is increasingly being emphasized in elementary school curricula.

If students' meanings for variables are inconsistent across idiosyncratic and informal representations of variables and/or task types or are founded in surface-structures that are not relevant when applying algebraic conventions to literal symbols, then students may struggle in coordinating their prior meanings with using a *specific* representation for variables (i.e., literal symbols) with a *specific* set of algebraic conventions. Yet, in discussions with mathematics education researchers following presentations of findings from this study at national and international conferences, a common assumption has been that student's conceptions for variables would be consistent across informal, idiosyncratic, and conventional representations of variables despite a lack of research supporting this assumption. How elementary students generalize and develop meaning for various symbolic representations of variables and what generalizations and meaning they have already developed, which is not well researched, is likely to contribute to our understanding of when and why students develop difficulties with literal symbolic representations of variables and algebraic conventions.

Hiebert and Carpenter (1992) described mathematics as being understood "if its mental representation is part of a network of representations" and that the "degree of understanding is determined by the number and strength of the connections" (p. 67). Drawing on Hiebert and Carpenter's description of understanding mathematics, the purpose of this study is to examine US grade 4-6 students' understanding of formal and informal representations of variables, as revealed in the types of numbers they substitute across various representations of variables and task types for addition problems with unknown addends prior to being introduced to algebraic conventions. Two research questions drove this part of the study:

1. What number substitutions do grade 4-6 students provide for various symbolic representations of unknown addends in dual unknown addend tasks?
2. What attributes of various symbolic representations of unknown addends do grade 4-6 students attend to in determining number substitutions for dual unknown addends tasks?

The author is not promoting or imposing a particular definition, symbolic representation, or conception of variable (e.g., Philipp, 1992; Usiskin, 1999). This study is attending to *student generated* mathematical generalization(s) and understanding(s) for various representations of unknown addends, regardless of how one defines, conceptualizes, or differentiates between variables, placeholders, unknowns, and generalized numbers in unknown addend tasks. The findings of this study supports further research into learning trajectories for variables (Blanton & Knuth, 2013; Marum et al., 2011), number and operation (Switzer, Buchheister, & Dougherty, 2014), and the transition to the highly symbolized and conventionalized subject of school algebra courses. In addition, the findings raise the question of whether the inclusion of informal and idiosyncratic representations for variables or if exclusive use of letters-as-variables, as used in the *Measure Up* program (Dougherty, 2004) and current research on learning trajectories for variables (Blanton & Knuth; Marum et al.), assists students' development of algebraic thinking and symbolization.

While not promoting a particular definition for variable, Philipp (1992) provided a definition for variable as "consisting of a symbol standing as a referent for a set consisting of at least two elements" (p. 557), which captures two important aspects that are particularly relevant to this study. When applying this definition,

> [e]ven the literal symbol $x$ in the statement $x + 3 = 7$ is a variable, because $x$ represents any of the elements of the set in the unstated but implicitly assumed domain, be it the real numbers, the rational numbers, the integers, the natural numbers, as so forth (p. 577).

First, Philipp distinguishes between symbols, or representations, used for the variable and the referent, set of numbers, for which the symbols stand. Second, and closely related to the first, is that the symbol that stands for the set does not have to be a literal symbol. Therefore, "variable" is understood broadly in this study to include a variety of *representations* and *referents* for varying quantities, placeholders, unknowns, and generalized numbers.

## Selective Review of Prior Research

Researchers have established that elementary grade students can engage in algebraic thinking and use literal symbol representations of variables to make generalizations and model situations (Carraher, Schliemann, & Schwartz, 2007; Dougherty, 2008; Soares, Blanton, & Kaput, 2006). Blanton and Knuth (2013) are developing a grade 3-7 learning progression for understanding variable that solely uses letter-symbolic representation of variable. Brizuela et. al., (2013) are developing a similar progression for grade K-2 students.

### Student difficulties with variables

Student difficulties with literal symbols have also been well documented (c.f., Booth, 1984; Ellis, 2007; Knuth, Alibali, McNeil, Weinberg, & Stephens, 2005; Küchemann, 1981; Swafford & Langrall, 2000; Warren & Cooper, 2007). For example, Fujii (2003) noted that students in middle grades and beyond often believe different letters must stand for different numbers (e.g., in $x + y = 12$, $x$ and $y$ cannot both be 6), which would be revealed in student's number substitutions. However, it is unknown whether elementary grade students hold this same belief for informal or idiosyncratic representations (e.g., in ☐+ = 12 or __ + __ = 12), formal algebraic representations (e.g., $x + y = 12$ or $y + y = 12$), or if this misconception arises later during instruction in an algebra course.

Research findings related to the types of numbers students substitute for variables represented with literal symbols have further revealed potential connections to students' conceptions of number and operation. Fujii and Stephens (2008) found that students often employ boundary values, such as addends cannot be greater than the sum. Similarly, Christou and Vosniadou (2012) reported findings consistent with their hypothesis that "students tend to consider literal symbols in algebra to stand for natural numbers only" (Christou, Vosniadou, & Vamvakoussi, 2007, p. 291). However, it is unknown whether elementary grade students also exhibit a natural number bias for informal or idiosyncratic representations or if this bias arises later during instruction in an algebra course.

**When and why difficulties with variables arise.** Research has established much about student difficulties for *conventional literal symbol* representations of variables *after* being introduced to algebraic conventions of use (e.g., Fujii & Stephens, 2008; Kaput, 2008; MacGregor & Stacey, 1997; Radford, 2000; Sfard & Linchevski, 1994). In fact, research on middle grade and higher students' conceptions and misconceptions of variable has exclusively focused on *conventional literal symbol* representations of variables (c.f., Fujii, 2003; Küchemann, 1981; Lucariello, Tine, & Ganley, 2014).

Corresponding research findings related to students' prior knowledge and experiences with commonly used *informal* representations of variables and unknowns commonly found in elementary textbooks, and when, why, or how documented student difficulties with literal symbol representations of variables arose are lacking. For example, Koedinger and Nathan (2004) found that secondary students demonstrated different meanings and strategies for dealing with variables that may vary across task types (i.e., word problems, word equations, and equations) and task difficulty. However, their findings did not take into consideration the different representations of variables within the tasks used in the study as potential factors contributing to these different meanings and strategies.

Early research related to elementary grade students' working with variables tended to involve situations where researcher(s) introduced students to letters as variables and told them what the letters meant, often resulting in initial student confusion or rejection of the notation (e.g., Booth, 1984; Carraher, Brizuela, & Schliemann, 2000; Stacey, 1989). For example, early in a study conducted by Carraher, et al., (2000) with third grade students learning in an "algebrafied" arithmetical setting, the researchers introduced *T* as a variable for a problem related to people's heights. When asked what the *T* stood for, two students, who had solved the general case, responded that the *T* stood for "tall" and "ten". The researcher explained that *T* could stand for whatever Tom's height might be, but the students were reluctant to accept this explanation. Over the next several classes, students began to use the words "whatever" and "any" to refer to the referent of the letters (e.g., *T* meant "whatever Tom's height is", or "any height"). However, the findings do not indicate if the students interpreted these letters as variables or if they were just adopting the researchers' terminology.

In addition to contextually understanding variables, such as Tom's height, students must also understand what numbers can be substituted for variables in different contexts and task types. Knuth, et al., (2005) suggested that middle school students' understanding of literal symbols might be fragile, particularly when representing *multiple-values*. They further suggested that the types of tasks in which elementary grade students engage, where literal symbols often have single-value solutions, might contribute to students' fragile understanding of literal symbols. In elementary grades, these values often consist of natural numbers, which may be a contributing factor to Christou and Vosniadou's (2012) findings that secondary students displayed a natural number bias when substituting values for variables.

**Variable surface structures.** The author hypothesizes that imposing any particular symbolic representation for variables can, and often will, result in unanticipated or unintended student interpretations. Brizuela, Gardiner, Sawrey, Newman-Owens, and Blanton (2013) reported how a first grade student who was told that *w* represented a quantity in a contextual problem stated that *w* was 21 because "w" was the twenty-first letter in the alphabet. In another situation, students were exploring the relationship between student heights and their height when wearing a one-foot tall hat. When asked how tall a boy who was *b* feet tall would be including his hat, a student stated the boy would be *c* feet tall because c was the next letter (one more than *b*) in the alphabet. These students appeared to apply the ordinality of letters in the alphabet as cardinality (e.g., b + 1 = c, where b and c are strictly letters).

Common misconceptions, such as those already noted, have been revealed in the extant research for literal symbols. Misconceptions such as those reported by Brizuela, et al. (2013) are derived from attributes of the specific symbolic representation used. Less is known about misconceptions that could arise due to students primarily attending to surface structures (Skemp, 1982) of informal or idiosyncratic representations of variables. Findings related to student difficulties with conventional letter-symbolic representations of variables are well established but less is known about student generalizations for commonly used informal representations of variables before and during the transition to algebra courses and how they may contribute to the research findings for literal symbols, algebraic conventions of use, including when and why difficulties and misconceptions arise.

## Representations of Variables as Source of Meaning

While algebraic conventions use literal symbols to represent variables, Kaput (2008) incorporated a broader view of symbolization including representations that are written, verbalized, and drawn. When applying Kaput's view of symbolization to variables, the extant research is lacking and further research warranted considering that teachers, mathematics education researchers, and curricular materials do not always represent variables using letters, especially in elementary grades.

The author reviewed three elementary mathematics textbook series (*enVisionMATH*, 2009; *Investigations in Number, Data, and Space*, 2008; *Mathematics: The Path to Math Success!*, 1998) and found commonly used representations of variables included, but were not limited to, blanks, shapes, words, and question marks. While algebraic conventions exist for the use of letter-symbolic representations of variables, no such conventions were included for these informal representations. Even if teachers decided to introduce the algebraic conventions for these informal representations of variables, there is no evidence that students would adopt these conventions across representations. In addition, introducing the algebraic conventions does not address the question of whether the use of various informal representations of variables helps or hinders students' future understanding and work with literal symbols. Therefore, the limited research on students' conception(s) of these informal representations of variables is problematic.

Carpenter, Franke, and Levi (2003) noted potential difficulties may arise from common representations of variables used in elementary grades like "*Find the different numbers you can put in the boxes:* ☐+ ☐= 9" (p. 75). They noted that using two boxes, the same representation for the variables, could be confusing to students and contribute to the development of misconceptions about the use of variables. For instance, in the number sentence, ☐+ ☐= 9, it is unclear if students would determine that the boxes have to represent the same number, since they are the same representation, or if they can represent different numbers. While conventions exist for the use of $x$ in $x + x = 9$, no conventions exist, or at least has not been clearly agreed upon, for "boxes" unless we retrospectively apply algebraic conventions to these informal representations.

The phrasing of the prompt for ☐+ ☐= 9 imposes a specific meaning on the representation of the boxes that the students may not hold. By asking students' what numbers, plural, they "*can put in the box*"; the box may be interpreted as a placeholder to be *filled* with various numbers. Therefore, this phrasing of the prompt may bias students not to interpret the boxes as a representation that *stands for* or *represents* numbers thereby hindering students from considering whether the use of the same representation has any relevance to the numbers that can be substituted for the boxes. This phrasing places the emphasis on a particular interpretation of the representation. Asking students what numbers the boxes could stand for, represent, or would satisfy the equation, and why, more closely aligns with and captures the essence of the conventional algebraic use of literal symbols as variables.

In recognizing the potential student difficulties that may arise from ☐+ ☐= 9, Carpenter, Franke, and Levi (2003) suggested that using the number sentence ☐+ Δ = 9 would be preferable. However, the suggestion that ☐+ Δ = 9 would be preferable assumes students would attend to and recognize the symbols ☐and Δ as different variables, the ☐'s in the original equation as representing the same variable, and a need for representing variables differently, which lacks support from research. In addition, in ☐+ Δ = 9 we do not know if elementary students would interpret this notation as *requiring* the ☐and Δ to be different values (i.e., they cannot be the same value) because they are different representations or if they would even attend to the representations being different.

Another common representation for unknowns, or variables, found in elementary grades mathematics textbooks is blanks (e.g., __ + __ = 12), which are uniquely problematic in representing variables. For instance, the equation ☐+ ☐= 9 could be rewritten as __ + __ = 9, maintaining the same representation for each unknown addend. However, rewriting the equation

☐+ Δ = 9 using blanks is problematic since blanks are generally not represented in different ways, making it difficult to maintain the different representations for the unknown addends. Therefore, the use of blanks may mask or hinder students' generalization of the convention that the same representation for variable in the same equations must be the same value and different representations for variables can have different values.

Within the extant research on variable, "variable" is often used to describe quantities that vary (referent) and the symbol used to represent varying or unknown quantities. Those of us engaged in research related to algebra and early algebra, have reified the *representation* of variables and are able to fluidly move back and forth between looking at and looking through these symbols (Kaput, 2000, 2008). Students must be provided with opportunities to develop fluency in looking at and looking through symbolic representations of variables, regardless of representation, and further research is needed to shed light on student generalizations for idiosyncratic and informal representations of variables. Since little is known about students' meanings for these various representations of variables, or how these may contribute to or hinder student understanding of literal symbols, the extant research does not provide an adequate foundation upon which to determine if the continued use of informal representations or solely using literal symbols would provide stronger support for students understanding of variable.

## Methods

Thirty-six grade 4-6 participants engaged in two semi-structured task based interviews to investigate their conception(s) for informal and formal symbolic representations of unknown addends. The core mathematical tasks consisted of addition tasks with two unknown addends and a known sum. Core mathematical tasks were written as two task types, equations with unknown addends represented as shapes, letters, or blanks, and word problems with the unknown addends represented with words. The interview protocol was purposefully developed to explore students' conceptions for representations of unknown addends *prior* to their introduction to the algebraic conventions for variables that can be compared to findings for students who have been introduced to algebraic conventions.

**Participants**

Thirty-six grade 4-6 students enrolled in one elementary or middle school from each of two research sites participated in this study. Participants' mathematics teachers rated each student for whom consent to participate was received as high, typically, or underachieving in mathematics based on their perceptions of each student. Students from each grade level and category were randomly selected to participate in the study. Participants from research site A, a Midwestern public school district, consisted of nine fourth graders, nine fifth graders, and six sixth graders[1]. Participants from research site B, a Southern public school district, consisted of one high, one low, and two typically achieving students per grade level.

**Task Design Framework**

Drawing on a review of the literature and commonly used representations of variables, place holders, unknown and task types found in US elementary grades mathematics textbooks, the task design framework, see figure 1, was developed and used to generate the tasks used in this study. Interview tasks were developed from five core mathematical tasks that systematically varied two dimensions or factors: 3 task types (word problems, word equations, and equations/expressions) X 4 representations of unknown addends (blanks, letters, shapes, and words) resulting in 12 possible tasks for each core mathematical task. Of the twelve possible tasks, those selected for inclusion in this study reflected the types in which US grade 4-6 students

[1] Due to a video camera malfunction during one sixth grader's interview, the number of students included in the results section is thirty-five.

would likely have previous experience. As shown by the highlighted components in figure 1, the included tasks consisted of word problems and word equations with the unknown addends represented by words and equations with the unknown addend represented with blanks, letters, or shapes.

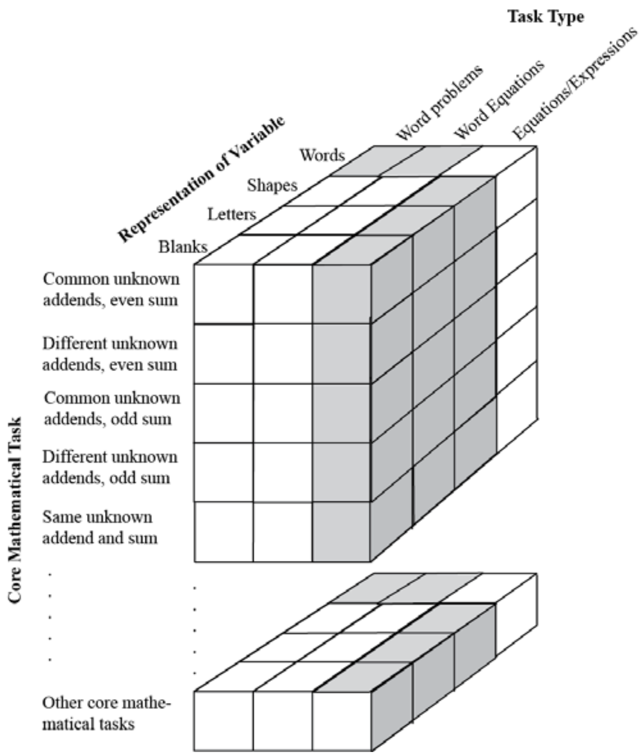


*Figure 1.* Task design framework.

## Interview Tasks

Interview one tasks, see figure 2, were generated from two core mathematical tasks - common or different unknown addends, with an even sum of 12, limiting the independent variables to the representations of the unknown addends and task type. The researcher hypothesized that participants would not discriminate equations with unknown addends represented with the same symbols from equations with the unknown addends represented with different symbols; but participants would distinguish between the same and different unknown addends in word problems by drawing on the context of the word problems.

| Core Mathematical Task | Order Presented | Task Type | Unknown Addend Representation | Task |
|---|---|---|---|---|
| Common unknown addends, even sum | 1 | WP | Word | Shakira and Tim have the same number of gummy bears. Together they have 12 gummy bears. How many gummy bears could Shakira have? How many gummy bears could Tim have? |
| | 4 | EQ | Blank | __ + __ = 12 |
| | 6 | EQ | Letter | $y + y = 12$ |
| | 2 | EQ | Shape | ☐+ ☐= 12 |
| Different unknown addends, even sum | 7 | WP | Word | Together Tom and Anne have 12 feet of ribbon. How long could Tom's ribbon be? How long could Anne's ribbon be? |
| | 3 | EQ | Letter | $a + b = 12$ |
| | 5 | EQ | Diff | ☐+ Δ = 12 |

*Figure 2.* Interview one unknown addend task by task type and representation of unknown addend.

**Semi-structured Interview Protocol**

Participants were shown a printed copy of each task in the same order and asked what numbers the unknown addends could be (e.g., "what could the [blanks, letters, or shapes] be?" or "how many gummy bears could Shakira have?"). Participants only providing a few number substitutions were asked if there were any other numbers the unknown addends could be. After providing their initial number substitutions participants were asked follow-up questions including, but not limited to,

- Are there any numbers that [one of the unknown addends] could not be,
- Can one of the unknown addends be [a given number greater than the sum], and
- Can one of the unknown addends [be a given rational number].

The follow-up questions were designed to gather further evidence of participants' conception(s) of the various representations of unknown addends including (a) initial and follow-up number substitutions, (b) providing multiple, single or no number substitutions for individual addends, (c) attending to and/or distinguishing between the same and different representations of unknown addends, and (d) providing the same, different, or same and different number substitutions for the unknown addends. Participants' responses were coded to indicate the set of numbers each participant included in their initial and follow-up number substitutions. The codes included natural {1, 2, 3,…}, whole {0, 1, 2, 3,…}, integer {…, -2, -1, 0, 1, 2, …}, and rational $\{\frac{a}{b}$, where $a$ and $b$ are integers and $b \neq 0\}$ for initial and follow-up number substitutions.

While natural, whole, and integer numbers are a subset of the rational numbers, responses were only coded as natural, whole, or integer number substitutions if the number substitutions did not include a fraction. Likewise, number substitutions that included any number in fraction form (e.g., $\frac{1}{2}, \frac{5}{3}, \frac{4}{4}$, or $\frac{-3}{3}$), was coded as a rational number substitution. In order to check coding reliability, two external researchers viewed and coded the first and second interview videos for three randomly selected students. Agreement between the researcher's and two external researcher's codes was 91.9%.

These codes were not intended to indicate if all possible numbers from the identified set were included. For example, a participant who stated that the *y*'s in *y* + *y* = 12 could be 6 and 6, 2 and 10, and 8 and 4, would be coded as providing a natural number substitution. If they had included 0 and 12 then the response would have been coded as a whole number substitution. If the participant included any integer or rational number substitutions, their response would have been coded as an integer and initial rational number substitutions respectively.

Participant responses were then coded for three response categories. First, participant *solutions* were coded to indicate how many number substitutions they provided. Responses were coded as *multiple substitutions* if they provided more than one pair of number substitutions, *single substitution* if they provided one number substitution, *no solution* if they indicated that there were no number substitutions that would work, and *no response/unsure* if no number substitutions were provided or if the participant was unsure what numbers could be substituted.

Second, participant responses were coded for *symbolic differentiation* based on participants' number substitutions, explanations, and responses to interview questions. *Symbolic differentiation* describes whether participants differentiated, mathematically or otherwise, between symbolic mathematical representations. In this study, symbolic differentiation reflected whether participants interpreted the unknown addends as representing the same or different quantities based on their number substitutions. In conventional algebraic terms, a response that the $y$'s in $y + y = 12$ can only be six or that the $x$ and $y$ in $x + y = 12$ have to be different values would indicate that the participant is distinguishing between the unknown addends. Participants providing multiple solutions, including six, for the $x$'s and $y$'s in the previous equations would be an indication that they did not distinguish between the representations for the unknown addends.

Finally, participant responses were coded for *symbolic reciprocity* based on the number substitutions provided for the pair of unknown addends in each task. Responses were coded as *same* if only the same number was substituted, *different* if only different numbers were substituted, and *both* if the same and different number substitutions were provided. If the participant did not provide a number substitution for the pair of unknown addends or stated that they were unsure, their response was coded as *No Response/Unsure.*

## Results

The following section presents findings for participants' initial number substitutions from the first interview, which is further disaggregated by grade level and task types, revealing the types of number substitutions participants provided for various symbolic representations of unknown addends. Next, findings for participants' follow-up number substitutions are presented and further disaggregated by grade level and task types. Finally, findings related to participants' response categories are reported, providing details into the attributes of the representations for the unknown addends participants attended to in determining their number substitutions.

**Participants' Initial Number Substitutions**

As shown in Table 1, participants' initial number substitutions consisted solely of natural (72.7%) and whole number (27.3%) substitutions. In addition, participants provided initial natural number substitutions more often for all tasks, ranging from 51.4% for the equation ☐+ Δ = 12 to 100% for the Gummy Bear (same unknown addend) word problem. Thirty-two number substitutions were provided for the equation ☐+ Δ = 12 as three participants were unsure what numbers could be substituted.

Table 1
*Initial Number Substitutions for each Interview Task.*

| **Task** | **Initial Natural Number Substitutions** | **Initial Whole Number Substitutions** |
|---|---|---|
| Shakira and Tim have the same number of gummy bears. Together they have 12 gummy bears. How many gummy bears could Shakira have? How many gummy bears could Tim have? | 35 (100%) | 0 (0.0%) |
| ☐+ ☐= 12 | 31 (88.6%) | 4 (11.4%) |
| $a + b = 12$ | 23 (65.7%) | 12 (34.3%) |
| __ + __ = 12 | 22 (62.9%) | 13 (37.1%) |
| ☐+ Δ = 12 | 18 (51.4%) | 14 (40.0%) |
| $y + y = 12$ | 22 (62.9%) | 13 (37.1%) |
| Together Tom and Anne have 12 feet of ribbon. How long could Tom's ribbon be? How long could Anne's ribbon be? | 25 (71.4%) | 10 (28.6%) |
| Total | 176 (72.7%) | 66 (27.3%) |

As shown in Table 2, the percentage of initial natural number substitutions provided by participants at each grade level for word problems was greater than the percentage provided for equations. Participants attended to the context of the word problems in determining their number substitutions, while the equations had no such context. The differences in grade level initial number substitutions for word problems and equations were not significantly different ($\chi^2(3) = 0.307$, $p < 0.80$) indicating that there was no statistical difference in the proportion of initial and whole number substitutions provided for word problems and equations across grade levels.

Table 2
*Initial Natural Number Substitutions by Task Type and Grade Level.*

| **Task Type** | **4th** | **5th** | **6th** | **All Participants** |
|---|---|---|---|---|
| Word Problems | 23 (88.5%) | 24 (92.3%) | 13 (72.2%) | 60 (86.1%) |
| Equations | 49 (75.4%) | 45 (69.2%) | 22 (48.9%) | 116 (63.9%) |

While 72.7% of participants provided initial natural number substitutions, seven participants, two fourth graders, one fifth grader, and four sixth graders, provided initial whole number substitutions more often than natural number substitutions, see Table 3. For the Gummy Bear word problem, all participants, including these seven students, indicated that Shakira and Tim would each have six Gummy Bears. For the Ribbon word problem, four of these seven participants indicated that Tom or Anne could have all, or none, of the ribbon. The other three participants provided number substitutions where Tom and Anne each had ribbon but did not explicitly state that Tom or Anne could *not* have all, or none, of the ribbon.

As can be seen in table 3, the seven participants provided three, six, seven, seven, and seven whole number substitutions for the five equation tasks respectively. Except for the equation $\square + \square = 12$, these seven participants were highly consistent in providing whole number substitutions for all equations and drew on the context of the word problems to decide between natural and whole number substitutions. The lower number of initial whole number substitutions for $\square + \square = 12$, or high percentage of initial whole number substitutions for the remaining questions, may have been a result of the participants applying whole number substitutions for the follow-up questions to subsequent initial number substitutions. However, there was no evidence from the interviews to support or refute this hypothesis.

Table 3
*Initial Number Substitutions for Participants Providing More Initial Whole Number Substitutions.*

| **Task** | **Initial Whole Number Substitutions** | **Initial Natural Number Substitutions** |
|---|---|---|
| Shakira and Tim have the same number of gummy bears. Together they have 12 gummy bears. How many gummy bears could Shakira have? How many gummy bears could Tim have? | 0 | 7 |
| ☐+ ☐= 12 | 3 | 4 |
| *a* + *b* = 12 | 6 | 1 |
| __ + __ = 12 | 7 | 0 |
| ☐+ Δ = 12 | 7 | 0 |
| *y* + *y* = 12 | 7 | 0 |
| Together Tom and Anne have 12 feet of ribbon. How long could Tom's ribbon be? How long could Anne's ribbon be? | 4 | 3 |
| Total | 33 | 15 |

**Word Problems.** The difference in the proportion of grade level participants' initial natural number substitutions for the two word problems, see Table 4, were not statistically significant ($\chi^2(2) = 0.837$, $p < 0.50$), indicating that grade-level differences did not result in differences in the proportion of initial natural and whole number substitutions for each word problem. All participants provided natural number substitutions, six Gummy Bears each, for the Gummy Bear word problem. For the Ribbon problem participants indicated that Tom and Anne could have different lengths of ribbon, multiple solutions. While participants provided multiple number substitutions for the ribbon problem, all participants exhibited quasi-variable thinking (Fujii, 2003)by not providing fractional feet of ribbon (e.g., Tom has 5 ½ feet of ribbon and Anne has 6 ½ feet of ribbon).

Table 4
*Initial Natural Number Substitutions for Word Problems by Grade Level.*

| **Word Problem** | **4th** | **5th** | **6th** |
|---|---|---|---|
| Gummy Bears | 13 (100%) | 13 (100%) | 9 (100%) |
| Ribbons | 10 (76.9%) | 11 (84.6%) | 4 (44.4%) |

**Equations.** The difference in the proportion of grade level participants' initial natural number substitutions for equations, see Table 5, were not statistically significant ($\chi^2(8) = 3.6465$, $p < 0.90$), indicating that grade-level differences did not result in differences in the proportion of initial natural and whole number substitutions for the different equations. Participants provided initial natural number substitutions for often for all equations except for __ + __ = 12 by sixth graders (33.3%) and ☐+ Δ = 12 by fifth graders (46.2%) and sixth graders (22.2%).

Table 5
*Initial Natural Number Substitutions for Equations by Grade Level.*

| **Task** | **4$^{th}$** | **5$^{th}$** | **6$^{th}$** |
|---|---|---|---|
| ☐+ ☐= 12 | 12 (92.3%) | 12 (92.3%) | 7 (77.8%) |
| *a* + *b* = 12 | 9 (69.2%) | 9 (69.2%) | 5 (55.6%) |
| __ + __ = 12 | 8 (61.5%) | 11(84.6%) | 3 (33.3%) |
| ☐+ Δ = 12 | 10 (76.9%) | 6 (46.2%) | 2 (22.2%) |
| *y* + *y* = 12 | 10 (76.9%) | 7 (53.8%) | 5 (55.5%) |
| Total | 49 (75.4%) | 45 (69.2%) | 22 (48.9%) |

Overall, the percentage of initial natural number substitutions decreased from 4$^{th}$ through 6$^{th}$ grade with 75.4%, 69.2%, and 48.9% provided respectively. The only individual equation where the percentage of initial natural number substitutions decreased from 4$^{th}$ through 6$^{th}$ grade was ☐+ Δ = 12 with 76.9%, 46.2%, and 22.2% respectively. The percentage of initial natural number substitutions provided by fourth and fifth graders were the same for ☐+ ☐= 12 (92.3%) and *a* + *b* = 12 (69.2%) and greater than those provided by sixth graders with 77.8% and 55.6% respectively. Fourth graders provided the greatest percentage of initial natural number substitutions for ☐+ Δ = 12 (76.9%) and $y + y = 12$ (76.9%). Fifth graders provided the greatest percentage of initial natural number substitutions for ___ + ___ = 12 (84.6%).

Christou and Vosniadou (2012) found that algebra students demonstrated a natural number bias in their number substitutions for literal symbols. The findings from participants' initial number substitutions from this study tend to support and extend Christou and Vosniadou's finding to fourth through sixth graders number substitutions for literal and other commonly used symbolic representations of unknown addends.

**Follow-up Number Substitutions**

After providing their initial number substitutions, each participants' thinking was further probed by asking follow-up questions including, but not limited to,

- Are there any numbers that [one of the unknown addends] could not be?
- Can one of the unknown addends be [a given number greater than the sum]?
- Can one of the unknown addends be [a given rational number]?

When asked if any numbers would *not* work for one of the unknown addends, participants consistently employed "boundary values" (Fujii & Stephens, 2008), stating that the unknown addends could not be greater than the sum. Participants who stated that the unknown addend could not equal or be greater than the sum were asked to explain their thinking. The following interaction with Brian, a sixth grader, for the task ☐+ ☐= 12 illustrates and is representative of participants' responses.

**I:** What numbers could the squares be?

**P:** Six, six plus six equals twelve, or it could be ten and two, or eleven and one.

**I:** So, are there any numbers that first square cannot be?

**P:** Any number above twelve.

**I:** Why can't it be above twelve?

**P:** Because, or it could be twelve because that [pointing at second ❑] could just be a zero. It couldn't be above twelve because it has to equal twelve.

**I:** So if this [pointing at first ❑] was thirteen what would the problem be?

**P:** The answer would be higher than twelve. It would be like… if it was thirteen and zero it would be thirteen.

The above interaction with Brian represents the thinking provided by participants who employed the sum as a boundary value. Such responses revealed the generalization that when adding two numbers the resulting sum would be greater than, or equal to, the greater addend. Ashlock (2006) noted that elementary grade students often mistakenly believe that adding makes numbers greater (e.g., 5 + 7 = 12, however -5 + 7 = 2). The fifth and sixth graders in this study had been introduced to integers and rational numbers in their mathematics classes, yet continued to apply generalizations for addition that were only true for natural and whole numbers (Switzer et al., 2014).

As shown in Table 6 the percentage of follow-up natural (29.8%) and whole number substitutions (69.0%) were nearly the same as the percentage of initial whole (27.3%) and natural number substitutions (72.7%) respectively, a net change of 103 (58.5%) initial natural number substitutions were coded as follow-up whole number substitutions. Participants continued to provide natural and whole number substitutions for the follow-up questions and not include rational or integer number substitutions.

Table 6
*Follow-Up Number Substitutions for each Interview Task.*

| **Task** | **Natural Number Substitutions** | **Whole Number Substitutions** |
|---|---|---|
| Shakira and Tim have the same number of gummy bears. Together they have 12 gummy bears. How many gummy | 35 (100%) | 0 (0.0%) |

| | | |
|---|---|---|
| bears could Shakira have? How many gummy bears could Tim have? | | |
| □+ □= 12 | 1 (2.9%) | 34 (97.1%) |
| *a* + *b* = 12 | 3 (8.6%) | 32 (91.4%) |
| __ + __ = 12 | 0 (0.0%) | 35 (100%) |
| □+ Δ = 12 | 3 (8.6%) | 29 (82.9%) |
| *y* + *y* = 12 | 9 (25.7%) | 26 74.3%) |
| Together Tom and Anne have 12 feet of ribbon. How long could Tom's ribbon be? How long could Anne's ribbon be? | 22 (62.9%) | 13 (37.1%) |
| Total | 73 (29.8%) | 169 (69.0%) |

When categorized by grade level, the proportion of follow-up natural and whole number substitutions for word problems and equations were not significantly different ($\chi^2(2) = 4.1134$, $p < .2$) indicating that grade level did not differentiate between the proportion of follow-up number substitutions for the word problems and equations included in this study. The number of initial number substitutions provided for word problems decreased by one at each grade level. The number of initial number substitutions provided for equations decreased by forty-three (87.8%), forty-two (93.3%), and fifteen (68.2%) respectively at each grade level.

Table 7
*Follow-up Natural Number Substitutions by Task Type and Grade Level.*

| **Task Type** | **4th** | **5th** | **6th** | **All Participants** |
|---|---|---|---|---|
| Word Problems | 22 (84.6%) | 23 (88.5%) | 12 (66.7%) | 57 (81.4%) |
| Equations | 6 (9.2%) | 3 (4.6%) | 7 (15.6%) | 16 (9.1%) |

While there was no statistically significant difference in the proportion of follow-up number substitutions for word problems and equations across grade levels, there was a statistically significant difference between the proportion of initial and follow-up natural number substitutions for word problems and equations ($\chi^2(1) = 40.0874$, $p < .001$). The decrease in the percentage of initial to follow-up natural number substitutions for word problems was 4.7 percentage points and 54.8 percentage points for equations.

**Word Problems.** The difference in the proportion of grade level participants' follow-up natural number substitutions for the two word problems, see Table 8, were not statistically significant ($\chi^2(2) = 1.217$ , $p < 0.70$), indicating that grade-level differences did not result in differences in the number of follow-up natural and whole number substitutions for each word problem. The only change in participants' initial and follow-up natural number substitutions was one participant at each grade level who indicated that Tom or Anne could have all, or none, of the ribbon for their follow-up number substitutions but not for their initial number substitutions.

Table 8
*Follow-up Natural Number Substitutions for Word Problems and Grade Level.*

| **Word Problem** | **4th** | **5th** | **6th** | **All Participants** |
|---|---|---|---|---|
| Gummy Bears | 13 (100%) | 13 (100%) | 9 (100%) | 35 (100%) |
| Ribbons | 9 (69.2%) | 10 (76.9%) | 3 (33.3%) | 22 (62.9%) |

**Equations**. As shown in table 9, for equations there was a net change of 100 initial natural number substitutions to follow-up whole number substitutions. The difference in the proportion of natural and whole number between initial and follow-up number substitutions for equations was statistically significant ($\chi^2(1) = 122.9274$, $p < .001$). Participants provided three or fewer follow-up natural number substitutions for each equation except $y + y = 12$, where they provided ten follow-up natural number substitutions. The difference in the proportion of follow-up natural and whole number substitutions for $y + y = 12$ and the remaining equations was significantly different ($\chi^2(1) = 17.9095$ $p < .001$).

When the three participants who indicated that the *y*'s had to be the same value (natural number substitution) were excluded from the analysis, the difference in the proportion of follow-up natural and whole number substitutions for $y + y = 12$ and the remaining equations was still significantly different ($\chi^2(1) = 9.5414$ $p < .01$). Participants responses to the initial and follow-up questions did not provide any evidence as to why the number of follow-up natural number substitutions for $y + y = 12$ were significantly greater than the remaining equations.

Table 9
*Follow-Up Number Substitutions for each Interview Equation.*

| **Task** | **Natural Number Substitutions** | **Whole Number Substitutions** |
|---|---|---|
| ☐+ ☐= 12 | 1 (2.9%) | 34 (97.1%) |
| $a + b = 12$ | 3 (8.6%) | 32 (91.4%) |
| __ + __ = 12 | 0 (0.0%) | 35 (100%) |
| ☐+ Δ = 12 | 3 (8.6%) | 29 (82.9%) |
| $y + y = 12$ | 9 (25.7%) | 26 74.3%) |
| Total | 16 (9.14%) | 156 (89.1%) |

**Summary.** Participants provided evidence of attending to the context of the word problems in determining the types of number substitutions. For the Gummy Bear problem, all students provided the natural number substitution of six Gummy Bears per person. For the Ribbon problem, participants provided natural and whole number substitutions with their interpretation of whether one of the people could have all, or none, of the ribbon was possible.

For equations, participants demonstrated evidence of a natural number bias in their number substitutions for equations. The natural number bias decreased from fourth through sixth grade with 75.4%, 69.2% and 48.9% initial natural number substitutions respectfully. In addition, while not statistically different, the evidence of a natural number bias was inconsistent across grade level and task types.

Overall, participants provided initial natural number substitutions more often than initial whole number substitutions for all tasks. When disaggregated by grade, participants provided initial natural number substitutions more often than initial whole number substitutions with the exception of fifth graders number substitutions for the tasks __ + __ = 12, ☐+ Δ = 12, and $y + y = 12$ and sixth graders who provided the same number of initial natural and whole number substitutions for ☐+ Δ = 12.

In contrast to participants' initial number substitutions, and Christou and Vosniadou's (2012) findings, participants' follow-up number substitutions revealed evidence of a whole number bias for the unknown addend equations and employed the sum as a boundary value in determining the type of number substitution. Yet, participants' initial and follow-up number substitutions for the word problems showed little change indicating that the participants' attention to the context of the word problems was consistent for initial and follow-up number substitutions.

Overall, participants did not attend to surface-structures (Skemp, 1982) of the unknown addends in equations. Three participants did attend to surface-structures of unknown addends represented with shapes and three attended to deep-structures of the $y$'s in $y + y = 12$ in stating that only six could be substituted. Furthermore, other than the three participants who attended to attributes of the shapes, there was no evidence that any participants believed that unknown addends represented with different symbols must stand for different numbers (Fujii, 2003).

Carpenter, Franke, and Levi (2003) noted potential difficulties that may arise from representations of variables commonly used in elementary grades, such as $\square + \square = 9$, and suggested that $\square + \Delta = 9$ would be preferable. However, the findings from this study indicate that the participants would provide similar number substitutions for $\square + \square = 9$ and $\square + \Delta = 9$. Further, using these shapes as representations of the variables was found to be confounding for the three participants who believed the number substitutions for the $\square$ should be greater than the number substitutions for the $\Delta$.

While these findings provide initial evidence of an overall follow-up whole number bias for equations and follow-up natural number bias for word problems, the proportion of follow-up number substitution types varied across task type and equations with the unknown addends represented with letters, and grade-levels. Therefore, there is initial evidence that the representation of the unknown addends, task type, and grade-level may have influenced the types of number substitutions provided by participants.

**Response Categories**

In the following section, findings from participants' number substitutions are further described in relation to the three participant response categories of *solutions, symbolic reciprocity,* and *symbolic differentiation.* Next, the three response categories are synthesized into a single *combined response category* that further reveal what participants' number substitutions may reveal about how they interpreted the unknown addends when represented with different symbolic representations and in different task types.

**Solutions.** As shown in table 10**,** all participants tended to provide multiple solutions, although not a full range of solutions, for the unknown addends, with the exception of the Gummy Bear problem (common unknown addends) where all participants indicated that Tim and Shakira would have six Gummy Bears each (single solution). For the Ribbon problem, each of the three participants providing a single solution for the unknown addends misinterpreted the problem, stating that Tom and Anne had to have the same amount of ribbon. For the equation $y + y = 12$, each of the six participants, four 4$^{th}$ graders and two 6$^{th}$ graders, stated that the values for the *y*'s *had* to be the same because the unknown addends were the both *y*'s. However, none of these participants applied the same reasoning to the equations ☐+ ☐= 12 or __ + __ = 12.

Table 10
*Solutions for each Interview Task.*

| Task | Multiple | Single | No Response |
|---|---|---|---|
| Shakira and Tim have the same number of gummy bears. Together they have 12 gummy bears. How many gummy bears could Shakira have? How many gummy bears could Tim have? | 0 | 35 | 0 |

| | | | |
|---|---|---|---|
| ☐+ ☐= 12 | 35 | 0 | 0 |
| a + *b* = 12 | 35 | 0 | 0 |
| __ + __ = 12 | 35 | 0 | 0 |
| ☐+ Δ = 12 | 33 | 0 | 2 |
| y + y = 12 | 29 | 6 | 0 |
| Together Tom and Anne have 12 feet of ribbon. How long could Tom's ribbon be? How long could Anne's ribbon be? | 32 | 3 | 0 |

**Symbolic Reciprocity.** In order to determine if participants differentiated between the symbolic representations of the unknown addends, responses were coded for symbolic reciprocity. In order to differentiate between participants who interpreted different representations of the unknown addends as having different number substitutions from those who believed the number substitutions *must* be different for each unknown addend, participant responses were coded as providing the same, different, different and same, or no response.

As shown in Table 11, with the exception of the Gummy Bear problem, participants tended to provide the same (i.e., 6 and 6) and different (e.g., 5 and 7) number substitutions for each unknown addend. Since providing the same number substitution only occurred when participants provided the same number substitution to the pair of unknown addends exclusively, the findings for "Same" in Table 11 are the same as those for "Single Solution" in Table 10.

Table 11
*Symbolic Reciprocity for each Interview Task.*

| **Task** | **Same** | **Different** | **Same and Different** | **No Response** |
|---|---|---|---|---|
| Shakira and Tim have the same number of gummy bears. Together they have 12 gummy bears. How many gummy bears could Shakira have? How many gummy bears could Tim have? | 35 | 0 | 0 | 0 |
| ☐+ ☐= 12 | 0 | 0 | 35 | 0 |
| a + *b* = 12 | 0 | 0 | 35 | 0 |
| __ + __ = 12 | 0 | 0 | 35 | 0 |

| | | | | |
|---|---|---|---|---|
| ☐+ Δ = 12 | 0 | 2 | 31 | 2 |
| y + y = 12 | 6 | 0 | 29 | 0 |
| Together Tom and Anne have 12 feet of ribbon. How long could Tom's ribbon be? How long could Anne's ribbon be? | 3 | 0 | 32 | 0 |

Of the 199 participants who provided multiple number substitutions, 197 (99.0%) provided the same and different number substitutions for the pair of unknown addends for each task. These findings provide evidence that the participants did not exhibit the misconception noted by Fujii (2003) that different variables *must* be different values. Of the tasks where the pair of unknown addends were the same (i.e., Gummy Bear problem, ☐+ ☐= 12, ___ + ___ = 12, and y + y = 12), all 199 multiple solution responses included the same and different number substitutions for each unknown addend. Of the tasks where different symbolic representations were used for the pair of unknown addends (i.e., Ribbon problem, a + *b* = 12, ☐+ Δ = 12), 98 of 100 responses included the same and different number substitutions for each unknown addend.

The two responses that did not include the same and different number substitutions for each unknown addend where the unknown addends were different occurred for ☐+ Δ = 12. In each case, the participant provided multiple solutions but indicated that the square and triangle could not be the same value. In addition, the two participants who did not provide a response for ☐+ Δ = 12 demonstrated similar thinking as the former participants, which will be described in the following section on symbolic differentiation.

**Symbolic Differentiation.** Responses were coded for symbolic differentiation to distinguish between participants who differentiate between tasks with different symbolic representations for the unknown addends and those who did not. As shown in Table 12, all participants differentiated between the unknown addends for the Gummy Bear problem and six participants differentiated between the *y*'s in *y* + *y* = 12 by recognizing that each unknown addend had to be the same value. The three participants who differentiated between the length of ribbon Tom and Anne had was a result of misinterpreting the problem, believing Tom and Anne had to have the same length of ribbon.

Table 12
*Symbolic Differentiation for each Interview Task.*

| **Task** | **Differentiated** | **Did not differentiate** |
|---|---|---|
| Shakira and Tim have the same number of gummy bears. Together they have 12 gummy bears. How many gummy bears could Shakira have? How many gummy bears could Tim have? | 35 | 0 |
| ☐+ ☐= 12 | 0 | 35 |
| a + *b* = 12 | 0 | 35 |
| __ + __ = 12 | 0 | 35 |
| ☐+ Δ = 12 | 4 | 31 |
| y + y = 12 | 6 | 29 |
| Together Tom and Anne have 12 feet of ribbon. How long could Tom's ribbon be? How long could Anne's ribbon be? | 3 | 32 |

Fujii (2003) noted that students in middle grades and beyond often believe different letters must stand for different numbers (e.g., in $x + y = 12$, $x$ and $y$ cannot both be 6). Students who correctly apply algebraic conventions and students who exhibit the misconception noted by Fujii would likely both provide multiple number substitutions for the $x$ and $y$ in $x + y = 12$. In addition, students exhibiting this misconception and students correctly applying algebraic conventions are both differentiating between the $x$ and $y$, by attending to the different representation to determine if the x and $y$ can be the same number or must be different numbers.

The number of responses coded as multiple solutions were also coded as not differentiating between unknown addends *except* for the equation ☐+ Δ = 12. The four participants who differentiated between the square and triangle consisted of two participants who did not provide a response and the two participants who provided only different number substitutions for the square and triangle. Each of these participants attended to surface structures (Skemp, 1982) of the square and triangle in determining their number substitutions. For instance, in the exchange below Trevor, a sixth grader, provided evidence of attending to the number of sides in attempting to determine his initial number substitution but was ultimately unsure how to interpret the shapes.

**I:** What about square plus triangle equals twelve?

**P**: I don't think we've done that. We might have but I can't remember. I'm thinking about [the state assessment] to see if there are any problems like that but I doubt it. I don't think we've learned the square plus triangle thing yet. I don't know what that means.

**I**: Did you have a guess of what you thought it meant?

**P**: I'd be guessing it's probably like the sides of this one (points at the ☐and begins counting the sides). One, two, three, four, five, six, seven, eight. I don't know what I'm thinking about.

Tom, a fifth grader, indicated the number of sides for each shape may determine the type of number substitutions. He indicated that since the square had four sides and the triangle had three sides the equation would make sense if it were multiplying the square and triangle. However, as demonstrated in the following exchange, he was unsure of how to make sense of these representations for the unknown addends in the context of addition.

**I**: What about that one [shows ☐+ Δ = 12]

**P**: I never saw that in class.

**I**: Square plus triangle equals twelve. What could the square and triangle be?

**P**: Um, it could be, because square has four sides you could multiply it (Δ) and get twelve but it wants plus. If that (+) wasn't there. I don't see how I could get twelve. I don't see how I could get twelve so much.

As with Trevor and Tom, Bill, a sixth grade student, attended to the geometric attributes of the square and the triangle in determining his number substitutions. He hypothesized that the relative size of the shapes in comparison to each other played a factor in determining the numbers that could be substituted for the square and triangle, as demonstrated in the following exchange.

**I**: What about this one, square plus triangle equals twelve? What could the square and triangle be?

**P**: It could be ten [❑] plus two [Δ] or eleven [pointing at ❑] plus one [pointing at Δ]. This [pointing at Δ] is like numbers like only with one number. This triangle is too little. Like six over here [Δ] and six here [❑] too but ten plus two will equal twelve.

**I**: Okay, so this [Δ) is only like a one-digit number?

**P**: Yea, one-digit number and this one [❑] two.

**I**: Does this one [❑] have to be a two-digit number?

**P**: No.

**I**: So this one [❑] could be a one digit?

**P**: Yea

**I**: But this one has to be a one digit?

**P**: Yea, because it's little.

Like Bill, Joy, a fourth grader, attended to the relative-size of the shapes. However, she makes a clearer distinction between the size of the shapes and the numbers that they can represent. Based on her reasoning she is also unsure if each shape could be six. Based on her response to the initial question this exchange was coded as an initial natural number substitution.

> **I**: So what about that one? For square plus triangle equals twelve, what could the square and triangle be?
>
> **P**: Maybe this [☐] can be the big number and this [Δ] the small number because the triangle is small and that [☐] is big. Like eleven plus one, and you can do three, you could put nine here [☐] and put a three there [Δ] because you can split it up, split it up into a part to be equal. You could put a four there [☐] plus an eight right there [Δ].

In each of these cases, the students drew on attributes of the square and triangle as important indicators of the types of numbers that could be substituted for each representation.

**Combined response categories.** Table 13 shows the composite response categories (i.e., solutions, symbolic reciprocity, and symbolic differentiation for each of the seven interview tasks. Participant responses for each task primarily fall into one of two composite response categories. For the Gummy Bear word problem (same unknown addend), all but one response provided a single solution where the same value was substituted for both unknown addends, which were interpreted as being the same. For three of the six remaining tasks (i.e., ☐+ ☐= 12, a + b = 12, and __ + __ = 12), all participants provided multiple solutions that consisted of the same and different number substitutions for the unknown addends, which were not differentiated between.

Table 13
*Combined Response Categories for each Interview Task.*

| **Solutions** | **NR** | **Single** | **Multiple** | | | |
|---|---|---|---|---|---|---|
| **Symbolic Reciprocity** | **NR** | **Same** | **Diff** | **Both** | | |
| **Symbolic Differentiation** | **Yes** | **Yes** | **Yes** | **NR** | **Yes** | **No** |
| Shakira and Tim have the same number of gummy bears. Together they have 12 gummy bears. How many gummy | 0 | 35 | 0 | 0 | 0 | 0 |

| | | | | | | |
|---|---|---|---|---|---|---|
| bears could Shakira have? How many gummy bears could Tim have? | | | | | | |
| ☐+ ☐= 12 | 0 | 0 | 0 | 0 | 0 | 35 |
| a + *b* = 12 | 0 | 0 | 0 | 0 | 0 | 35 |
| __ + __ = 12 | 0 | 0 | 0 | 0 | 0 | 35 |
| ☐+ Δ = 12 | 2 | 0 | 2 | 1 | 0 | 30 |
| y + y = 12 | 0 | 6 | 0 | 0 | 0 | 29 |
| Together Tom and Anne have 12 feet of ribbon. How long could Tom's ribbon be? How long could Anne's ribbon be? | 0 | 3 | 0 | 0 | 0 | 32 |

Algebraic conventions of use for the same literal symbols in unknown addend equations consist of (a) a single *solution*, (b) the same value substituted for the same literal symbols (*symbolic reciprocity)*, and (c) understanding that the same surface structure (Skemp, 1982) of the literal symbols has mathematical consequences for what numbers may, and may not, be substituted (*symbolic differentiation*). As previously reported, of the interview equations with the unknown addends represented with the same symbolic representation, only six participants combined response categories were consistent with a conventional algebraic solution. Each of the six participants indicated that the *y*'s in *y* + *y* =12 had to both be 6 and only 6. None of the six participants applied their combined response categories for *y* + *y* =12 to other equations with the same symbolic representation for the unknown addends (i.e., ☐+ ☐= 12 or __ + __ = 12).

For the three equations with the same symbolic representation of the unknown addends, ninety-nine (94.3%) of participants combined response categories consisted of multiple *solutions*, *symbolic reciprocity* consisting of the same and different number substitutions, and no *symbolic differentiation.* All seventy responses reflected this combined response category for the equations ☐+ ☐= 12 and __ + __ = 12.

Algebraic conventions of use for different literal symbols in unknown addend equations consist of (a) multiple *solutions*, (b) the same and different values substituted for the same literal symbols (*symbolic reciprocity*), and (c) understanding that the different surface structure (Skemp, 1982) of the literal symbols have mathematical consequences for what numbers may, and may not, be substituted (*symbolic differentiation*). For the two interview equations with different symbolic representations for the unknown addends ($a + b = 12$ and □+ Δ = 12), none of the participants' combined response categories corresponded with a conventional algebraic solution.

Even though sixty-five (92.9%) participants' number substitutions matched number substitutions generated from a conventional algebraic interpretation of the unknown addends with the exception of participants quasi-variable thinking (i.e., excluding rational and integer number substitutions), participants showed evidence of not considering the different symbolic representations to be mathematically important. With the exception of the six participants who stated that the *y*'s in $y + y$ =12 had to each be six and the five participants who attended to the surface structures of the □and Δ in determining number substitutions for □+ Δ = 12, participants provided the same number substitutions for equations where the unknown addends were the same or different for letters or shapes (i.e., $a + b$ =12 and $y + y$ =12, and □+ □= 12 and □+ Δ = 12).

## Discussion

While tasks such as those in this study included symbolic representations for the unknown addends that are commonly found in elementary mathematics classes, results indicate students' number substitutions alone are insufficient in determining *how* students determine their number substitutions and how their interpretation of the symbolic representation for the unknown addends contribute to their selection of number substitutions. Teachers and researchers must also consider how participants attend to surface structures of the symbolic representations of the variables, unknown addends in this study, and what deep structures they have generalized, if any, including *symbolic reciprocity* and *symbolic differentiation*.

Two participants did not provide a solution, both for the equation $\square + \Delta = 12$, by attending to surface structures (Skemp, 1982) and concluding that the triangle might represent "smaller" numbers and the square represent "bigger" numbers. Likewise, the six participants who provided a single solution (i.e., provided the same number substitution for each unknown addend), interpreted the surface structure being the same as meaning the unknown addends had to be the same value. However, these six participants only applied this generalization to the equation $y + y = 12$.

Categorizing multiple number substitution *solutions* for *symbolic reciprocity* and *symbolic differentiation* resulted in three distinct groups, all for equations. Of the three multiple solution composite categories, one included two participants providing only different number substitutions for $\square + \Delta = 12$ because the $\square$ represented "bigger" number than the $\Delta$, or the $\Delta$ represented "smaller" numbers than the $\square$. It is important to note that these participants did not provide only different solutions for the $\square$ and $\Delta$ because they were different symbolic representations but because they attributed the "smaller" triangle to "smaller" numbers and the "bigger" square to "bigger" numbers and did not provide only different number substitutions to the other equations with different unknown addends.

Eighty percent of all responses and 93.7% of equation responses were included in the final multiple number substitution *solutions* subgroup. These responses consisted of participants who provided multiple *solutions*, the same and different number substitutions for *symbolic reciprocity,* and did not exhibit *symbolic differentiation.* The only tasks where participants consistently differentiated between unknown addends were word problems where they could draw on the context to assign meaning to the unknown addends. For equations, where no corresponding context was provided, participants did not differentiate between the same and different symbolic representations of unknown addends with the noted exceptions for $y + y = 12$ and ☐+ Δ = 12.

Understanding that the same representation for variables in the same equation or expression must be the same value(s) and different representation for variables in the same equation or expression can be the same and/or different value(s), within the constraints of the problem, is a fundamental aspect of algebraic conventions that algebra students must understand. Algebra students must attend to and differentiate between the surface structures (Skemp, 1982) of sameness or differentness of the symbolic representations and then mathematically interpreting the deep structure of the symbolic representations.

While participants inclusion of only different number substitutions for the square and triangle in the equation ☐+ Δ = 12 could initially be inferred to indicate that they believe that different representations for the unknowns must be different values (Fujii, 2003), the inference would misdiagnose the participants thinking. Further, while the remaining responses to ☐+ Δ = 12 and all responses to equations a + *b* = 12 would parallel those of students correctly applying algebraic conventions, participants' responses that the *a* and *b or* ☐and Δ can take on the same and different values was not due to recognizing that these different symbols could mathematically be understood as representing all solutions including when the symbols were the same and different values. Instead, participants interpreted these equations in the same way as they interpreted equations with the same symbolic representations for the unknown addends in ☐ + ☐= 12, $y + y = 12$, and __ + __ = 12.

The findings in this study indicate that participants tended to not attend to surface structures, and thereby did not attend to deep structures, of the symbolic representations for the unknown addends in equations, with the noted exceptions for $y + y = 12$ and $\square + \Delta = 12$. In fact, participants' generalizations for making sense of the representations of the unknown addends were inconsistent with algebraic conventions of use and common misconceptions and difficulties demonstrated by algebra students (Booth, 1984; Knuth, Alibali, McNeil, Weinberg, & Stephens, 2005; Macgregor & Stacey, 1997; Swafford & Langrall, 2000; Warren & Cooper, 2008). The lack of findings in this study paralleling the difficulties and misconceptions for variables by algebra students suggests that these difficulties and misconceptions occur later, possibly during introduction to algebraic conventions in algebra classes. Further, the lack of strategies paralleling algebraic conventions of use may indicate that participants have not been provided with opportunities to consider whether different representations for unknowns is mathematically important or, if they have had such opportunities, they did not generalize them to the tasks in this study.

Finally, these findings also reflect participants' quasi-variable thinking (Fujii, 2003; Fujii & Stephens, 2008) and an initial natural number bias (Christou et al., 2007; Christou & Vosniadou, 2012). Participants regularly limited their initial number substitutions to natural numbers, applied the sum as a boundary value (Fujii & Stephens, 2008), and thereby did not attend to the full range a variation for the unknown addends. In addition, the findings for participants' follow-up number substitutions did not support Christou and Vosniadou's (2102) finding of a natural number bias. Participants initial natural number bias was not a good predictor of follow-up natural number bias. In fact, a higher percentage of participants provided follow-up whole number substitutions than initial natural number substitutions.

The author would like to acknowledge the contributions of Dr. Maria Blanton, Dr. Eric Knuth, Dr. Sarah Quebec-Fuentes, and the journal reviewers in providing feedback and suggestions from prior drafts of this manuscript. Their input was invaluable and made this a much stronger paper.

## References

Alibali, M. W., Nathan, M. J., Chiurch, R. B., Wolfgram, M. S., Kim, S., & Knuth, E. J. (2013). Teachers' gestures and speech in mathematics lessons: Forging common ground by resolving trouble spots. *Zdm, 45*, 425-440. doi:10.1007/s11858-012-0476-0

Ashlock, R. B. (2006). *Error Patterns in Computation*. Upper Saddle River, NJ: Pearson Education, Inc.

Blanton, M., & Knuth, E. (2013). Progression for developing understanding of variable across grades 3-7. *Project LEAP: Learning through an Early ALgebra Progression.* Retrieved from http://algebra.wceruw.org/documents/GA_LP_web.pdf

Booth, L. R. (1984). *Algebra: Children's Strategies and Errors*. Windsor, Berkshire: The NFER-NELSON Publishing Company.

Brizuela, B., Blanton, M., Gardiner, A., Gravel, B., & Newman-Owens, A. (2013). *Bringing variable notation to the forefront of early mathematics education*. Paper presented at the National Council of Teachers of Mathematics Research Pre-Session, Denver, CO.

Carpenter, T. P., Franke, M. L., & Levi, L. (2003). *Thinking Mathematically: Integrating Arithmetic and Algebra in Elementary School*. Portsmouth, NH: Hienemann.

Carraher, D., Schliemann, A., & Brizuela, B. (2000). *Early algebra, early arithmetic: Treating operations as functions*. Paper presented at the 22nd meeting of the Psychology of Mathematics Education, Tuscon, AZ. Available on CD-ROM retrieved from

Carraher, D. W., Brizuela, B., & Schliemann, A. (2000). *Bringing out the algebraic character of arithmetic: Instantiating variables in addition and subtraction*. Paper presented at the Psychology of Mathematics Education (PME), Hiroshima, Japan.

Carraher, D. W., Schliemann, A. D., & Schwartz, J. (2007). Early algebra is not the same as algebra early. In J. Kaput, D. W. Carraher, & M. Blanton (Eds.), *Algebra In the Early Grades* (pp. 235-272). New York, NT: Lawrence Erlbaum Associates.

Christou, K. P., & Vosniadou, S. (2012). What Kinds of Numbers Do Students Assign to Literal Symbols? Aspects of the Transition from Arithmetic to Algebra. *Mathematical Thinking and Learning, 14*(1), 1-27. doi:10.1080/10986065.2012.625074

Christou, K. P., Vosniadou, S., & Vamvakoussi, X. (2007). Students' interpretations of literal symbols in algebra *Reframing the Conceptual Change Approach in Learning and Instruction* (pp. 283-297): Emerald Group Publishing.

Dougherty, B. (2004). Early Algebra: perspectives and assumptions. *For The Learning of Mathematics, 24*(3), 28-30. Retrieved from http://www.jstor.org/stable/40248469 .

Dougherty, B. (2008). Measure Up: A quantitative view of early algebra. In J. J. Kaput, D. W. Carraher, & M. L. Blanton (Eds.), *Algebra in the Early Grades* (pp. 389-412). New York: Lawrence Erlbaum Associates.

Ellis, A. B. (2007). Connections between generalizing and justifying: Student's reasoning with linear relationships. *Journal for Research in Mathematics Education, 38*(3), 194-229.

*enVisionMATH*. (2009). (Vol. 2011): Pearson Education.

Fujii, T. (2003). *Probing students'understanding of variables through cognitive conflict problems: Is the concept of a variable so difficult for students to understand?* Paper presented at the 27th annual conference of the INternational Group for the Psychology of Mathematics Education, Honolulu, Hawaii.

Fujii, T., & Stephens, M. (2008). Using number sentences to introduce the idea of variable. In C. E. Greenes & R. Rubenstein (Eds.), *Algebra and Algebraic Thinking in School Mathematics: Seventieth Yearbook* (pp. 127-149). Reston, VA: National Council of Teachers of Mathematics.

Hiebert, J., & Carpenter, T. P. (1992). Learning and teaching with understanding. In D. A. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning* (pp. 65-100). Reston, VA: National Council of Teachers of Mathematics.

*Investigations in Number, Data, and Space*. (2008). (2 ed.). Glenview, Ill.: Pearson/Scott Foresman.

Kaput, J. J. (2000). Teaching and Learning a New Algebra With Understanding.

Kaput, J. J. (2008). Algebra from a symbolization point of view. In J. J. Kaput, D. Carraher, & M. L. Blanton (Eds.), *Algebra in the Early Grades* (pp. 19-25). New York: Lawrence Erlbaum Associates.

Knuth, E. J., Alibali, M. W., McNeil, N. M., Weinberg, A., & Stephens, A. C. (2005). Middle school students' understanding of core algebraic concepts: Equivalence & variable. *Zdm, 37*(1), 68-76.

Knuth, E. J., Stephens, A. C., McNeil, N. M., & Alibali, M. W. (2006). Does understanding the equal sign matter? Evidence from solving equations. *Journal for Research in Mathematics Education, 37*(4), 297-312.

Koedinger, K. R., & Nathan, M. J. (2004). The real story behind story problems: Effects of representations on quantitative reasoning. *Journal of the Learning Sciences, 13*(2), 129-164. doi:DOI 10.1207/s15327809jls1302_1

Küchemann, D. (1981). Children's Understanding of Mathematics. In K. Hart (Ed.), *Algebra* (pp. 11-16). London: Murray.

Lannin, J. K., van Garderen, D., Switzer, J. M., Buchheister, K., Hill, T., & Jackson, C. (2013). The mathematical development in number and operation for struggling first graders. *Investigations in Mathematics Learning, 6*(2), 19-47.

Lobato, J., Burns Ellis, A., & Muñoz, R. (2003). How "focusing phenomena" in the instructional environment support individual students' generalizations. *Mathematical Thinking and Learning, 5*(1), 1-36.

Lucariello, J., Tine, M. T., & Ganley, C. M. (2014). A formative assessment of students' algebraic variable misconceptions. *The Journal of Mathematical Behavior, 33*, 30-41. doi:10.1016/j.jmathb.2013.09.001

MacGregor, M., & Stacey, K. (1997). Students' understanding of algebraic notation: 11-15. *Educational Studies in Mathematics, 33*, 1-19.

Marum, T., Isler, I., Stephens, A. C., Gardiner, A., Blanton, M., & Knuth, E. (2011). *From specific value to variable: Developling students' avilities to represent unknowns.* Paper presented at the 33rd Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, Reno, NV.

*Mathematics: The Path to Math Success!* (1998). Silver Burdett Ginn.

McNeil, N. M., Grandau, L., Knuth, E. J., Alibali, M. W., Stephens, A. C., Hattikudur, S., & Krill, D. E. (2006). Middle-school students' understanding of the equal sign: The books they read can't help. *Cognition and Instruction, 24*(3), 367-385. doi:DOI 10.1207/s1532690xci2403_3

National Academy of Science. (2005). *Rising Above The Gathering Storm: Energizing and Employing America for a Brighter Economic Future* (Retrieved from http://www.nap.edu/catalog/11463.html). Washington, D.C.: National Acadamies Press.

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. Retrieved from Washington, DC:

Philipp, R. A. (1992). The many uses of algebraic variables. *The Mathematics Teacher, 85*(7), 557-561.

Powell, S. R. (2012). Equations and the Equal Sign in Elementary Mathematics Textbooks. *Elem Sch J, 112*(4), 627-648. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/22707760

Radford, L. (2000). Signs and meanings in students' emergent algebraic thinking: A semiotic analysis. *Educational Studies in Mathematics, 42*(237-268).

RAND Mathematics Study Panel, & Ball, D. L. (2003). *Mathematical Proficiency for All Students: Toward a Strategic Research and Development Program in Mathematics Education*. Santa Monica, CA: RAND Corporation.

Sfard, A., & Linchevski, L. (1994). The gains and the pitfalls of reification: The case of algebra. *Educational Studies in Mathematics, 26*(2/3), 191-228. Retrieved from http://www.jstor.org/stable/3482784

Skemp, R. R. (1982). Communicating mathematics: Surface structures and deep structures. *Visible Language, 16*(3), 281-288. Retrieved from <Go to ISI>://WOS:A1982PR32300008

Soares, J., Blanton, M. L., & Kaput, J. T. (2006). Thinking Algebraically across the elementary school curriculum. *Teaching Children Mathematics*, *12*(5), 228-235.

Stacey, K. (1989). Finding and using patterns in linear generalising problems. *Educational Studies in Mathematics*, *20*(2), 147-164. Retrieved from http://www.jstor.org/stable/3482495

Swafford, J. O., & Langrall, C. W. (2000). Grade 6 students' preinstructional use of equations to describe and represent problem situations. *Journal for Research in Mathematics Education, 31*(1), 89-112. doi:Doi 10.2307/749821

Switzer, J. M., Buchheister, K., & Dougherty, B. (2014). Examining number: A soft idea. *Teaching Children Mathematics, 21*(2), 69-73.

Usiskin, Z. (1999). Conceptions of school algebra and uses of variables. In B. Moses (Ed.), *Algebraic Thinking, Grades K-12: Readings from NCTM's School-based Journals and Other Publications* (pp. 7-13). Reston, VA: National Council of Teachers of Mathematics.

Warren, E., & Cooper, T. (2007). Generalising the pattern rule for visual growth patterns: Actions that support 8 year olds' thinking. *Educational Studies in Mathematics*, *67*(2), 171-185. doi:10.1007/s10649-007-9092-2