Contextual Cues To Word Learning: Mapping Meaning To Form At Two Developmental Milestones

Leslie E. Hodges
Georgia State University

Follow this and additional works at: http://scholarworks.gsu.edu/psych_theses

Recommended Citation
http://scholarworks.gsu.edu/psych_theses/136

This Thesis is brought to you for free and open access by the Department of Psychology at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Psychology Theses by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.
ABSTRACT

In two studies, we examined what constitutes a supportive context for word learning. In Study 1, we examined children’s comprehension of iconic gestures by asking 2, 3, and 4 year-old children, as well as adult controls, to choose the referent of an iconic gesture. The gesture either highlighted a possible action with the object, or a salient physical attribute of the object. By age 2, children performed above chance for iconics of action, but not for iconics of attribute; indicating that early use of iconic gesture to support word learning should utilize iconics of action. In Study 2, we examined how different levels of contextual support affect word learning from written contexts. Participants read very rare English words in contexts that were either: high (H), medium (M), or low (L) constraint. Participants had the greatest accuracy on a synonym judgment task when words were trained in a scaffolded sequence (H-M-M-L).

INDEX WORDS: Language development, Iconic gesture, Gesture comprehension, Assistance dilemma, Tier 2 words
CONTEXTUAL CUES TO WORD LEARNING: MAPPING MEANING TO FORM
AT TWO DEVELOPMENTAL MILESTONES

by

LESLIE HODGES

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of
Master of Developmental Psychology
in the College of Arts and Sciences
Georgia State University
2015
CONTEXTUAL CUES TO WORD LEARNING: MAPPING MEANING TO FORM
AT TWO DEVELOPMENTAL MILESTONES

by

LESLIE HODGES

Committee Co-Chair: Gwen Frishkoff
Committee Co-Chair: Şeyda Özçalışkan

Committee: Gwen Frishkoff
Şeyda Özçalışkan
Rebecca Williamson

Electronic Version Approved:

Office of Graduate Studies
College of Arts and Sciences
Georgia State University
May 2015
ACKNOWLEDGEMENTS

This thesis would not have been possible without the knowledgeable guidance of both of my advisors, Dr. Gwen Frishkoff and Dr. Şeyda Özçalışkan. I appreciate the flexibility and support that Dr. Rebecca Williamson provided throughout this thesis. I would like to thank Lauren J. Stites, Samantha N. Emerson, & Nevena Dimitrova for the invaluable support and feedback they provided. I would also like to thank Deborah Green, Gina Helms, Mark Vickers, and Lauren Schmuck who helped with data collection and coding. Thank you also to the participants and their families for their time. Finally, I would like to extend my thanks to the Language and Literacy Initiative for their funding support.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>ix</td>
</tr>
<tr>
<td><strong>1 GENERAL INTRODUCTION</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>2 STUDY 1: ROLE OF GESTURE IN WORD LEARNING</strong></td>
<td>3</td>
</tr>
<tr>
<td>2.1 Gesture Production and Word Learning</td>
<td>3</td>
</tr>
<tr>
<td>2.2 Gesture Comprehension and Word Learning</td>
<td>6</td>
</tr>
<tr>
<td>2.3 Overview of Study 1</td>
<td>9</td>
</tr>
<tr>
<td><strong>3 STUDY 1 METHOD</strong></td>
<td>10</td>
</tr>
<tr>
<td>3.1 Participants</td>
<td>10</td>
</tr>
<tr>
<td>3.2 Data Collection Procedure</td>
<td>10</td>
</tr>
<tr>
<td>3.2.1 Gesture comprehension task: Warm-up</td>
<td>10</td>
</tr>
<tr>
<td>3.2.2 Gesture comprehension task: Test trials</td>
<td>11</td>
</tr>
<tr>
<td>3.2.3 Object familiarity task</td>
<td>13</td>
</tr>
<tr>
<td>3.3 Data Analysis Procedure</td>
<td>13</td>
</tr>
<tr>
<td><strong>4 STUDY 1 RESULTS</strong></td>
<td>14</td>
</tr>
<tr>
<td><strong>5 STUDY 1 DISCUSSION</strong></td>
<td>17</td>
</tr>
<tr>
<td><strong>6 STUDY 2: CONTEXTUAL LEARNING OF TIER 2 WORDS</strong></td>
<td>20</td>
</tr>
<tr>
<td>6.1 Tier 2 Word Learning</td>
<td>21</td>
</tr>
</tbody>
</table>
6.2 Contextual Support for Tier 2 Word Learning ........................................... 22
  6.2.1 Arguments for independent reading .................................................. 23
  6.2.2 Arguments for direct instruction ...................................................... 24
  6.2.3 An intelligent tutor for optimal CWL .............................................. 25
6.3 Scaffolding of Support in CWL ............................................................... 26
  6.3.1 Scaffolding of contextual support .................................................... 29
6.4 Overview of Study 2 .............................................................................. 30

7 STUDY 2 METHOD .................................................................................... 32
  7.1 Participants .......................................................................................... 32
  7.2 Experiment Stimuli .............................................................................. 33
    7.2.1 Target words .................................................................................. 33
    7.2.2 Training contexts .......................................................................... 34
  7.3 Vocabulary assessments ....................................................................... 35
    7.3.1 Familiarity task ............................................................................. 36
    7.3.2 Synonym task ................................................................................ 37
  7.4 Data Collection Procedure .................................................................. 38
    7.4.1 Session 1: Pretest .......................................................................... 39
    7.4.2 Session 2: Training + Immediate posttest .................................. 39
    7.4.3 Session 3: Delayed posttest ......................................................... 40
  7.5 Procedure for Data Analysis ................................................................. 41
LIST OF TABLES

Table 1. *Gestures used in the comprehension task*. ...................................................... 13

Table 2. *Properties of target and non-target words* ....................................................... 34

Table 3. *Context constraint examples* ........................................................................... 35

Table 4. *Mean proportion correct on the Synonym Task* .............................................. 42
LIST OF FIGURES

Figure 1. Sample picture-pair for the warm-up forced choice task. .................. 11

Figure 2. Sample picture-pair for the test trials of the forced choice task ........... 12

Figure 3. Mean iconic gesture comprehension ........................................... 16

Figure 4. Familiarity ratings by word type at pretest and delayed posttest .......... 37

Figure 5. Overview of contextual word learning study .................................. 39

Figure 6. Synonym task scores ..................................................................... 43

Figure 7. Difference scores as measures of learning and forgetting .................. 44
1 GENERAL INTRODUCTION

Vocabulary development is critical to academic success largely because vocabulary skills are highly correlated with reading comprehension scores (Wagner, Muse, & Tannenbaum, 2007). Importantly, most words are learned in everyday contexts, such as reading and conversation, rather than through direct instruction (Sternberg, 1987). For example, when presented with the sentence “The weather was oragious, so we couldn’t play outside,” the reader can infer that ‘oragious’ means something like stormy or rainy, based on linguistic cues (e.g., the conjunction ‘so,’ which implies a causal link between ‘oragious’ and ‘couldn't play outside’), as well as real-world knowledge (people don't like to play outside when the weather is bad). Similarly, in conversation, speakers can provide contextual cues to meaning through gestures that either reinforce or supplement linguistic cues to meaning. This is especially important during the early (pre-literate) stages of word learning. For example, if a mother says to her child “Let’s play with the basketball” while holding her cupped hands slightly apart in the shape of a basketball, the child can infer that basketball refers to a round object. In both cases, the context — whether written, or spoken with gesture — can provide important cues to word meaning and can therefore support vocabulary development.

The present project addresses two central aspects of contextual word learning: comprehension of iconic gestures as a nonverbal symbol for a referent (Study 1), and the use of written contexts to learn a verbal symbol for a referent —specifically literary, or Tier 2, words (Study 2). In Study 1, we ask how early children begin to identify the referent for a nonverbal iconic symbol and whether the type of iconicity plays a role in this process. We compare iconic gestures highlighting action (moving hand as if bouncing a basketball) to iconic gestures highlighting attributes (holding hands apart to indicate shape of a ball). Prior work shows that deictic (pointing)
gestures facilitate early word learning. For example, a child is more likely to learn the word ‘book’ if an adult points to a book while saying the word ‘book’ (Tan & Schafer, 2005). By contrast, little is known about the role of iconic gestures in vocabulary development. Study 1 is a first step towards addressing this gap of in our knowledge concerning of iconic gesture’s contribution to vocabulary development. We predict that by age 3, children will be able to identify that referent of an iconic gesture that conveys action or attribute information, and we also predict that comprehension for iconic gestures conveying actions will precede the comprehension of iconic gestures conveying attributes. This study will also serve as the basis for future work examining gesture as an effective context for novel word learning in pre-literate children.

In Study 2, we ask what features of written contexts make a difference for word learning and retention. We consider competing hypotheses about the effectiveness of high-constraint contexts (sentences with strong cues to meaning) versus medium-constraint contexts (sentences with weaker cues to meaning), and compare the added value of presenting multiple contexts that vary in constraint. Starting in late childhood, written language becomes increasingly important as a source for word learning. In middle elementary grades, teachers have often observed a ‘4th grade slump,’ that is, a precipitous decrease in reading comprehension, despite continued progress in math, science, and other subjects (Chall, 1983). A likely source of this drop in performance is weak knowledge of literary, or Tier 2, words (e.g., ‘abstain’, ‘portend’) which rarely occur in speech but are essential for understanding written text (Leach, Scarborough, Rescorla, 2003). Importantly, previous work has shown that learning Tier 2 words requires exposure to these words in multiple, diverse contexts, to highlight different aspects of meaning. However, we know surprisingly little about which types of contexts lead to robust word learning (Lampinen & Faries, 1994). Study 2 is a first step towards addressing this gap. We will examine contextual
word learning (CWL) outcomes among healthy adults; we have two specific hypotheses: (1) contexts that are moderately informative rather than highly informative will result in better long term knowledge retention (i.e., higher scores on a delayed post-test) due to increased learner engagement, and (2) exposure to a variety of increasingly less informative contexts will result in the highest levels of word knowledge retention since learners will be provided with enough information initially and progressively challenged. The results from this study will serve as the basis for future research on contextual word learning among older children and adolescents.

2 STUDY 1: ROLE OF GESTURE IN WORD LEARNING

Young children frequently gesture, and these gestures play a significant role in language acquisition. Importantly, different gesture types appear in children’s repertoire at different times, and there is some evidence that children’s comprehension of each gesture type co-occurs with their production of the same gesture type. Here, we focus on children at the pre-literate level and examine their ability to identify the referent of an iconic gesture—a key ability that predicts later vocabulary development (Acredolo & Goodwyn, 1988). In this study, we focus on the subtypes of one particular category of gesture, namely iconic gestures, which convey characteristic actions or attributes associated with objects. We ask whether children’s comprehension of iconic gestures mirrors the pattern observed in their production of such gestures—with earlier comprehension of iconic gestures conveying action information than those conveying attribute information.

2.1 Gesture Production and Word Learning

Children’s gesture production is tightly related to early language development. The gestures children produce can predict emerging mastery of language milestones such as first words and first sentences (Iverson & Goldin-Meadow, 2005). Gesture can also expand a child’s vocabulary beyond their repertoire in speech, and even indicate to an adult that they are ready for time-
ly input in order to learn new words in speech (Özçalıșkan & Goldin-Meadow, 2011; Goldin-Meadow, Goodrich, Sauer, & Iverson, 2007).

The majority (~80-90%) of the gestures children produce at the early stages of language learning consist of deictic gestures indicating objects (e.g., pointing at a bottle; Iverson, Capirci, Volterra, Goldin-Meadow, 2008; Özçalıșkan & Goldin-Meadow, 2005). These gestures emerge in children’s communicative repertoires somewhere between 10 and 12 months of age, primarily in contexts of joint attention with a caregiver around a set of familiar objects (Bates, 1976; Carpenter, Nagell, Tomasello, Butterworth & Moore, 1998). Importantly, these deictic gestures can reliably be used as predictors of children’s impending language development for early language milestones. Children routinely point at objects before they produce spoken labels for the same objects; in fact, pointing to refer to an object reliably predates a child’s first words by about 3 months (Iverson & Goldin-Meadow, 2005). In addition, the earlier a child points at a particular object, the earlier the same child produces a spoken label for that object (Iverson & Goldin-Meadow, 2005).

The production of these early deictic gestures not only predicts children’s emerging vocabularies in speech, the individual variability observed in early gesture production predicts later vocabulary outcomes. The greater number of deictic gestures a child produces at age 1;6, the larger their spoken vocabulary one year later (Özçalıșkan, Adamson, Dimitrova, under review). Similarly, the number of objects indicated through deictic gestures at age 1;2 can reliably predict children’s vocabulary size at 4 and 5 years of age (Rowe, Özçalıșkan, & Goldin-Meadow, 2008; Rowe & Goldin-Meadow, 2009). Gesture continues to serve as a precursor to linguistic change beyond the one-word speech production stage. Before two word combinations are produced, children use speech + gesture combinations, for example saying ‘eat’ while pointing to a cookie
to express a desire to eat a cookie (Iverson & Goldin-Meadow, 2005; Özçalışkan, & Goldin-Meadow, 2005; Iverson, Capirci, Volterra, & Goldin-Meadow, 2008). Used in this way, the deictic gesture serves to supplement the information the child is producing in speech.

Children’s production of deictic gestures also serves to create language-learning opportunities; by pointing at objects, children can elicit timely input of object labels from adults. These parental translations of children’s deictic gestures into words significantly increase the likelihood that the particular word will enter a child’s spoken vocabulary (Goldin-Meadow, Goodrich, Sauer, & Iverson, 2007; Özçalışkan & Dimitrova, 2013; Wu & Gros-Louis, 2014).

Unlike deictic gestures, which derive their meanings from the immediate communicative context, iconic gestures derive their meaning through their form, by either conveying actions (e.g., bouncing) or attributes (e.g., round shape) that are characteristic of objects. These gestures emerge somewhat later - several months after children begin to produce their first words and typically after producing similar relational terms such as verbs in speech (Özçalışkan, Gentner, & Goldin-Meadow, 2014). This means a child might begin using a verb like ‘bounce’ before producing a gesture that indicates a bouncing action. Apart from the iconic baby signs taught deliberately by parents (Acredolo & Goodwyn, 1985, 1988; Goodwyn, Acredolo & Brown, 2000), the incidence of spontaneous iconic gestures in children’s early communicative repertoires remains relatively low, accounting for 1-5% of their gestures (Nicoladis, Mayberry & Genesee, 1999; Özçalışkan & Goldin-Meadow, 2005, 2009, 2011).

The first reliable increase in children’s iconic gesture production is observed around 26 months of age, at which point children begin to convey a significantly greater range of relational meanings using a greater number of iconic gestures (Özçalışkan & Goldin-Meadow, 2011; Özçalışkan et al., 2014). Looking more closely at the increase in iconic gesture production occur-
ring around 26 months shows that children are more likely to produce a gesture indicating actions they performed with the object rather than the perceptual qualities, or attributes, of the object (Acredolo & Goodwyn, 1988; Özçalışkan & Goldin-Meadow, 2011). This means that children are more likely to produce a bouncing gesture (action) to indicate a ball rather than holding their hands apart to indicate the size and shape (attributes) of the ball. These early iconic gestures serve an important purpose in that they expand the repertoire of early verbs found in children’s speech (Özçalışkan et al., 2012).

In sum, children’s production of deictic gestures, but not iconic gestures, can predict vocabulary development milestones, such as the production of first words and first two-word combinations. Iconic gestures do not predict these milestones, but they do allow children to communicate with a wider set of meanings than they have accessible through speech (Iverson, Capirci, Volterra, Goldin-Meadow, 2008; Özçalışkan et al., 2012). In terms of word learning, children’s deictic gestures provide an opportunity for parents to provide timely verbal input, which in turn predicts vocabulary development.

2.2 Gesture Comprehension and Word Learning

Previous work suggests that children comprehend different gesture types around the same time they begin to produce them, both for deictic and iconic gestures. Beginning with deictic gestures, children show comprehension around the same time as initial production. For example, children are able to follow the trajectory of an adult’s pointing gesture to a visible referent by 11 months of age (Carpenter et al., 1998), and by 12 months they can comprehend the intent of a pointing gesture to indicate the location of a hidden object (Behne, Liszkowski, Carpenter, & Tomasello, 2012). In fact, children who pointed to help a naïve experimenter find a hidden object were more likely to comprehend when an experimenter pointed to help the child locate the
hidden object (Behne et al., 2012). In addition, at both pre-linguistic and one-word stages of language development, pointing scaffolds comprehension of a referent: children are less likely to misidentify the referent of a word if an adult points in addition to providing a spoken label for the referent as compared to only providing a spoken label (Zukow-Goldring, 1996). Not surprisingly, the number of deictic gestures a mother produces while interacting with her child at 16 months is correlated with the size of the child’s vocabulary at 20 months (Iverson, Capirci, Longobardi, & Caselli, 1998). The existing work thus marks children’s ability to produce and comprehend deictic gestures at the same age, roughly around age one.

Turning to iconic gesture comprehension, research shows that 26 months marks the age that children not only significantly increase production of iconic gestures, but also significantly increase their comprehension of iconic gestures (Namy, Campbell, & Tomasello, 2004; Özçalışkan, Gentner, & Goldin-Meadow, 2014). As shown by Namy and colleagues (Namy & Waxman, 1998; Namy, 2001; Namy, Campbell, & Tomasello, 2004), 18-month-olds display a lack of sensitivity to iconicity as evidenced by their equally likely tendency to associate an iconic gesture (e.g., hopping V-shaped fingers up and down to represent a hopping rabbit) or an arbitrary gesture (holding a hand shaped in an arbitrary form to represent a rabbit) with an object. By contrast, 26 month-olds are more likely to associate an iconic gesture than an arbitrary gesture with an object, showing increased sensitivity to iconicity in their comprehension of non-verbal symbols (Namy et al., 2004). As such, the existing work on iconic gestures—similar to deictic gestures—suggests co-emergence of comprehension and production for iconic gestures.

A recent study tested children’s ability to infer meaning from iconic gestures from gesture-speech combinations where some information was only in gesture and not in the accompanying speech (i.e., supplementary iconic gesture+ speech combinations, Stanfield, Williamson, &
Özçalışkan, 2013), so that the gesture was supplementary and vital to accurate comprehension. In the study, an experimenter would say ‘I’m eating this one’ + hands held apart as if holding sandwich moving towards mouth. The child was then asked what the experimenter was eating – a bowl of cereal (incorrect choice) or a sandwich (correct choice). The study found that children were above chance at inferring information found in iconic gestures by age 3 (Stanfield et al., 2013), roughly around the same time that children begin to produce iconic gestures (Özçalışkan & Goldin-Meadow, 2011).

Similar to deictic gestures, the provision of iconic gestures can support word learning. For example, Zammit & Schafer (2011) showed that a mother providing an iconic gesture along with a spoken object label increased children’s receptive vocabulary for that word. In fact, the more iconic gestures a mother produced, the larger the child’s receptive vocabulary (Zammit & Schafer, 2011). More specifically, McGregor, Rohlfing, Bean, Marschner (2009) showed that 2-year-olds demonstrated better comprehension of the spatial term ‘under’ when the word was trained in a speech+iconic gesture condition (as compared to speech only or a photo of the ‘under’ relationship). In addition, Goodrich & Kam (2009) showed that children aged 2 to 4 demonstrated comprehension of novel verbs after the meaning of the novel verb was only expressed through iconic gesture.

However, earlier work focused on iconic gestures as a whole, leaving the developmental trajectory children follow in acquiring comprehension of subtypes of iconicity relatively unexplored. Iconic gestures can convey information about objects in one of two fundamental ways: by depicting action associated with an object (i.e., iconics of action; e.g., flapping arms to convey bird) or by depicting physical features associated with an object, such as its size and shape (i.e., iconics of attribute; e.g., holding palm with extended fingers to convey bird; Özçalışkan &
Goldin-Meadow, 2011). Earlier work that examined the production of each type of iconic gesture shows an advantage for iconics of action, but does not address comprehension. Young children (ages 1-3) use a significantly greater number of iconic gestures conveying action than ones conveying attribute information (Iverson, Capirici & Caselli, 1994; Özçalışkan & Goldin-Meadow, 2011); they also on average produce their first iconic gesture conveying action earlier than iconics conveying attribute information (Özçalışkan, Gentner, & Goldin-Meadow, 2014). However, we do not yet know whether the comprehension of each iconic gesture subtype follows the pattern observed in their production—with children understanding iconics of action earlier than iconics of attribute.

2.3 Overview of Study 1

In Study 1, we ask (1) at what age do children comprehend the referent of an iconic gesture and (2) whether comprehension of iconic gestures highlighting action emerges earlier than comprehension of iconic gestures highlighting attribute. We predict that children’s early comprehension of iconic co-speech gestures can follow one of two paths: One possibility is that children will comprehend iconic co-speech gestures conveying action and attribute around the same time. Both types of iconics convey relational information about objects, and as such, might be equally difficult for children to understand at the early ages. An alternative possibility is that children would comprehend iconic co-speech gestures conveying action earlier than the ones conveying attribute, thus mirroring the patterns found in their production of each iconic gesture subtype (Özçalışkan & Goldin-Meadow, 2011; Özçalışkan et al., 2014). Overall, we expect a significant increase in accuracy with age. Considering that children produce more action-oriented
iconic gestures, and children’s production follows comprehension in speech and deictic gesture use, we hypothesize that younger children will make more correct inferences when provided with iconic gestures conveying action.

3 STUDY 1 METHOD

3.1 Participants

The sample consisted of 54 children at the ages 2 (M<sub>age</sub> = 2;8, 10 females), 3 (M<sub>age</sub> = 3;2, 7 females), and 4 (M<sub>age</sub> = 4;3, 7 females)—with 18 participants per age group, along with 18 adult native speakers of English (M<sub>age</sub> = 21;5, 13 females). All children were learning English as their native language, and all adults were college students.

3.2 Data Collection Procedure

A female experimenter interviewed each participant individually in a laboratory setting. All participants completed a warm-up task to familiarize them with the experimental procedure. Then, they completed two test tasks: a gesture comprehension and an object familiarity task.

3.2.1 Gesture comprehension task: Warm-up.

The experimenter and child were seated at a table facing each other. The experimenter introduced the task with the following instruction: “I have a lot of different toys. I’m going to use my hands to tell you what toy I have. Then I’m going to ask you to tell me what toy I have. It’s not always going to be the same toy, so you’re going to have to pay close attention. Let’s try one.” The experimenter then conducted two warm-up trials to familiarize the child with the task before continuing onto the test trials. The warm-up trials used a baseball and a rabbit. For each warm-up trial, the experimenter produced one iconic gesture along with neutral speech; the warm-up gestures conveyed both characteristic action and characteristic feature associated with an object so as not to bias the child to one iconic gesture subtype (e.g., “I have this one” + fin-
gers in the shape of bunny ears hopping from right to left; see Figure 1 for the picture pair in the warm-up forced-choice task). Every child who showed understanding of the task demands by completing both warm-up trials with clear choices continued with the test trials, regardless of their accuracy in making the choice.

![Figure 1. Sample picture-pair for the warm-up forced choice](image)

### 3.2.2 Gesture comprehension task: Test trials.

Each participant was shown 12 iconic gestures, accompanied by minimally informative speech; 6 of the gestures conveyed *action* information (e.g., ‘I have this one’+flapping arms as if flying) and the remaining 6 conveyed *attribute* information (e.g., ‘I have this one’+linking hands by thumbs with fingers outspread in the shape of a bird) associated with a referent (see Table 1). After each iconic gesture, the experimenter presented the child with pictures of two referents: a correct match (e.g., bird) and an incorrect match (e.g., basketball) for the referent depicted in gesture. The child was then asked to choose one of the pictures (‘Which one did I have?’). Each child was presented with both the action and the attribute gesture for each referent (see Figure 2 for sample picture pair in the forced-choice task in test trials). The participant indicated their choice either verbally or by pointing to the picture.

The forced-choice pictures were always presented in the same pairs (bird/ball, alligator/book, bracelet/toothbrush), but the presentation order of the pairs was counterbalanced across
participants in blocks, to eliminate possible order effects and ensure that the same object pairs were never presented twice in a row. For example, one participant saw bird/ball, alligator/book, bracelet/toothbrush (with that sequence repeated four times in order to see all of the gestures) while another participant saw alligator/book, bracelet/toothbrush, bird/ball. Each object in the picture-pair was correct 50% of the time. For each block, the gesture type (action or attribute) order was varied, though at no time were there 3 action or 3 attribute gestures presented in a row. In addition, the side (left or right) of the correct object choice was randomized to control for a possible side bias.

Figure 2. Sample picture-pair for the test trials of the forced choice task

Six common objects familiar to 2-to 4-year-old children were chosen as gesture referents. Each object was paired with both an iconic gesture highlighting salient action and an iconic gesture highlighting a salient attribute associated with that object. For example, for the referent ‘bird’ the action gesture was hands flapping like wings and the attribute gesture was hands spread with thumbs linked like wings (see Table 1 for a full list of referents and the associated gestures used in the study). We also controlled for gesture viewpoint, using only gestures that conveyed a characteristic action or feature of an object from an observer viewpoint (i.e., gestures that do not incorporate the gesturer’s body into the gesture space; McNeill, 1992).
Table 1. *Gestures used in the comprehension task.*

<table>
<thead>
<tr>
<th>Object</th>
<th>Attribute Gesture</th>
<th>Action Gesture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alligator</td>
<td>Forearms together in mouth shape</td>
<td>Flat palms open and closing like an alligator snout</td>
</tr>
<tr>
<td>Basketball</td>
<td>Hands indicating size and round shape</td>
<td>Right hand, index finger extended, small up and down</td>
</tr>
<tr>
<td></td>
<td></td>
<td>motion as if tracing path of ball</td>
</tr>
<tr>
<td>Bird</td>
<td>Hands together, extended as if wings</td>
<td>Hands apart, horizontally flat, flapping motion</td>
</tr>
<tr>
<td>Book</td>
<td>Open hands in shape of book</td>
<td>Motion of opening hands as if opening a book</td>
</tr>
<tr>
<td>Bracelet</td>
<td>Trace gesture of the jewelry</td>
<td>Cupped right hand moves over left as if putting on</td>
</tr>
<tr>
<td></td>
<td></td>
<td>bracelet</td>
</tr>
<tr>
<td>Toothbrush</td>
<td>Horizontal extended right forearm by face</td>
<td>Extended index finger, by face, moving left to right</td>
</tr>
</tbody>
</table>

3.2.3 *Object familiarity task.*

Upon completion of the gesture comprehension task, each child completed an object familiarity task. In this task, child was presented with a picture of each of the six objects, and asked to label it. This task was intended to ensure that differences in children’s responses were not due to a lack of familiarity with the objects.

3.3 *Data Analysis Procedure*

For the iconic gesture comprehension task, participants’ responses to each of the 12 object pairs in the forced-choice gesture comprehension task received a score of ‘0’ (incorrect) or ‘1’ (correct). Scores were tabulated for each child separately for iconic gestures conveying action (score range = 0-6) and iconic gestures conveying attribute (score range = 0-6). The arcsine-transformed comprehension scores were then analyzed using a mixed ANOVA, with age (2, 3, 4, adult) as a between-subjects factor and iconic gesture type (action, attribute) as a within-subject
factor. We also compared children’s comprehension score against chance with a set of independent t-tests, separately by age and by gesture type.

For the object familiarity task, we analyzed children’s responses at labeling objects with a mixed ANOVA, with age (2, 3, 4) as a between-subjects factor, and object (alligator, basketball, bird, book, bracelet, toothbrush) as a within-subject factor.

### 4 STUDY 1 RESULTS

Children’s comprehension of iconic co-speech gestures of both types improved with age, $F(3, 68) = 27.64, p < .001, \eta^2 = .55$. Post-hoc tests with the Bonferroni correction showed no significant increase in overall comprehension scores between ages 2 and 3 ($M = 3.8, SD = 1.2$ vs. $M = 4.19, SD = 1.09$ respectively, $ns$), but significant increases between ages 3 and 4 ($M = 4.19, SD = 1.09$ vs. $M = 5.11, SD = .81$ respectively, $p = .02$). Adults were also significantly better than the 4-year-olds ($M = 5.11, SD = .81$ vs. $M = 5.97, SD = .11$ respectively, $p = .001$), suggesting continued growth of iconic gesture comprehension at the later ages.

Importantly, comprehension varied by iconic co-speech gesture type, showing a main effect of gesture type, $F(1, 68) = 6.16, p = .02, \eta^2 = .08$, but no interaction, $F(3, 68) = .87, p = .46$. Overall, participants showed better comprehension of iconics conveying action than iconics conveying attribute information. Given our focus on the early development of iconics of action versus iconics of attribute, we used planned comparisons to probe the performance of two- and three-year-old children by iconic gesture type. The two-year-olds’ comprehension scores were significantly better for iconics of action than iconics of attribute ($M_{\text{action}} = 4.17, SD = 1.33$ vs. $M_{\text{attribute}} = 3.44, SD = 1.38, t(17) = 2.77, p = .01$). However, comparing three-year-olds’ comprehension of each gesture type showed no significant difference between the two subtypes of iconic gestures ($M_{\text{action}} = 4.23, SD = 1.49$ vs. $M_{\text{attribute}} = 4.11, SD = 1.02; t(17) = .96, p = .35$). This
pattern was also reflected in the individual children’s responses: 61% of 2-year-olds (11/18) showed better comprehension of iconics conveying action than iconics conveying attribute, while 78% of 3-year-olds (14/18) showed similar comprehension (score difference ≤ 1) of iconics conveying action and attribute information.

We also compared children’s performance in the gesture comprehension task to chance levels (50% or 3/6). As can be seen in Figure 3, these results also showed differences by gesture subtype. Children were able to identify the referent of an iconic co-speech gesture conveying action significantly above chance levels by age 2, $t(17) = 3.69, p < .01, d = 1.79$ —a pattern that remained unchanged at the later ages. However, it was only at age 3 that children showed above-chance performance identifying the referent of an iconic co-speech gesture conveying attribute information, $t(17) = 4.61, p < .01, d = 2.24$.

We then examined whether children were familiar with the objects used in the study by analyzing the responses provided in the object familiarity task. This was to make sure that differences in comprehension were not a result of differences in familiarity with the objects. Responses were considered correct if they indicated knowledge of the object with words (e.g., ‘tweet tweet,’ ‘birdie,’ ‘seagull,’) or with gestures (e.g., flapping arms as if a bird). For this analysis, several participants were excluded: five two-year-olds for refusing to provide any response for more than half of the items, and one four-year old for a lack of videotaped responses. All of the age groups performed well on this task – with a mean accuracy above 80% across items. Unsurprisingly, there was a main effect of age on the number of correct responses, $F(2, 45) = 6.07, p = .005, \eta^2_p = .21$. However, post-hoc comparisons using the Bonferroni correction showed no difference between 2- and 3-year-olds’ proportion of correct responses ($M = .87, SD = .04$ vs. $M = .82, SD = .03$ respectively, $ns$), though 4-year-olds provided significantly more correct responses.
than 3-year-olds ($M = .82, SD = .03$ vs. $M = .97, SD = .03$ respectively, $p = .004$). Thus, it is unlikely that the comprehension differences observed between the 2- and 3-year-olds were a result of differences in familiarity with the items. There was also a main effect of object, $F(5, 225) = 8.80, p < .001, \eta^2 = .16$, but no significant interaction of object and age. The most correct responses were provided for ‘basketball’ ($M = 1.00, SE = .00$) and the least correct responses were provided for ‘bracelet’ ($M = .67, SE = .03$).

**Figure 3.** Mean iconic gesture comprehension.
Scores for iconic gestures conveying action (dark bars) or attribute (light bars) associated with objects by age (maximum possible score for each iconic gesture type = 6; error bars represent standard error).
5 STUDY 1 DISCUSSION

In this study, we asked whether children’s comprehension of iconic co-speech gestures go hand-in-hand with their production of such gestures. We found evidence that this is true, with the comprehension of iconics conveying action information emerging earlier than the comprehension of iconics conveying attribute information. Our results showed that children understood iconics of action reliably above chance around age 2—at least six months earlier than they showed comprehension of iconic gestures conveying attribute information. Children also steadily improved their comprehension of iconic gestures of both types with age, and were equally good at understanding both types of iconic co-speech gestures by age 3.

Why do children understand iconic gestures conveying action earlier than the ones conveying attribute information? One possible explanation is that comprehension of each iconic gesture type is largely driven by their production frequency. As shown in a recent longitudinal study (Özçalişkan et al., 2014; Özçalişkan & Goldin-Meadow, 2011) that followed the gestures and speech produced by 40 children from 14 to 34 months of age, children on average produced their first iconics of action by age 25 months. In contrast, the onset age for the production of iconics conveying attribute was later, roughly around 30 months. Even by 34 months, the last observation period for the study, the number of children producing iconics of action was higher than the ones producing iconics of attribute (38/40 vs. 28/40). Not only were iconics of action produced earlier, they were also produced more frequently, accounting for 76% of all iconic gestures children produced between ages 14 to 34 months. Previous work has shown strong correlations between the production and comprehension of deictic gestures (e.g., Behne et al., 2012). Our results extend this possible link to iconic gestures by showing that children show earlier comprehension of those iconic gesture types that they also produce earlier and with greater frequency.
Another possible explanation is that iconics of action provide more information than iconics of attribute. For example, the iconic action gesture ‘FLYING’, which involves flapping extended arms on sides of the body, conveys both the action of flying and the attribute of wing shape, and as such might be easier to understand than an iconic gesture conveying attribute (e.g., fingers outspread with thumbs connected to represent wings). The additional information may support children’s comprehension by giving multiple types of information within the gesture itself, thus providing more cues to meaning of the gesture for the child.

A third possible explanation is that iconics of action might be easier for children because of their close alignment with bodily experience. From a young age, children perform symbolic actions on objects, for example, picking up a toy phone and pretending to talk into it. This action can be decontextualized and performed with a substitute object representing the phone and eventually produced fully as a gesture (Iverson, 2010). This developmental progression of decontextualizing actions with objects may in turn result in earlier comprehension of iconics of action.

It is important to address the underlying assumption in classifying these iconic gestures as iconics of action and iconics of attribute – that the iconic gesture is interpreted as referring to the characteristics of the object and not representing the object itself. This is particularly important for the iconics of attribute – for example, with the basketball gesture used in this study it is not possible to differentiate if the gesture actually refers to ‘ball’ or if it refers to the adjective-like meaning of ‘round’. A future study, using a similar setup to Study 1, could tease these two possibilities apart. An experimenter could produce the same gesture for basketball – two hands held apart to indicate the size and shape of a basketball – but the forced choice options would include two balls - one matching in size and shape (basketball) and one much smaller but matching in shape (baseball). If what we are calling ‘iconics of attribute’ are indicating relational
meanings rather than object meanings, children should choose the option matching one in size (or other physical characteristics in other trials), otherwise, children should be equally likely to pick either of the ball options.

In addition to addressing the issue of iconics of attribute, future work could address the use of iconic gestures to support word learning. Goodrich and Kam (2009) used iconic action gestures to successfully teach novel verbs to 2-4 year olds, as measured by a forced choice-task. If iconics of attribute are expressing relational meaning, these gestures could be used to teach novel adjectives to 3-4 year olds. Since 2 year olds did not indicate comprehension of iconics of attribute above chance in Study 1, there would be little reason to expect word learning from iconics of attribute at that age.

In sum, Study 1 showed that comprehension emerges earlier for iconics of action than for iconics of attribute—a pattern that is also evident in the production of each iconic gesture subtype (Özçalışkan et al., 2014). This study extends earlier results with deictic gesture to the domain of iconic gesture by showing that comprehension and production of each gesture type follow a developmental pattern that goes hand-in-hand. This has important implications for the purposeful use of iconic gesture to support young children’s abstract word learning.
6  STUDY 2: CONTEXTUAL LEARNING OF TIER 2 WORDS

Broad and robust knowledge of words is important for the development of reading comprehension (NRP, 2000): fluent decoding, background knowledge, and skilled inferencing cannot substitute for understanding key words in a text (Cromley & Azevedo, 2007). In fact, studies estimate that a reader must already know between 90 and 95% of the words in a text to comprehend the meaning (Nagy & Scott, 2000).

The importance of vocabulary for reading comprehension increases in Grades 3–5, as students make the transition from learning-to-read to reading-to-learn (Chall, 1983; Stanovich, 1986). At this stage, readers are confronted with more challenging texts and are expected to recognize and comprehend more challenging ‘academic’ words (Beck, McKeown, & Kucan, 2013). Beck and associates have grouped words into three general categories: Tier 1 words, which are common and often occur in speech (‘smart’, ‘loner’, ‘take’), Tier 2 words which are abstract, polysemous, and rarely occur in speech (‘shrewd’, ‘recluse’, ‘abscond’), and finally, Tier 3 words which are content area specific (‘isotope’, ‘chiffonade’, ‘iambic’) (Beck, McKeown, & Kucan, 2002). In this study, we focus on Tier 2 words because these are found across academic subjects, for example, a socioeconomic ‘hypothesis’ could be discussed in a social studies class, and an experimental ‘hypothesis’ could be discussed in a science class. Because Tier 2 words refer to commonly shared concepts in a nuanced way, they are ideal to target for instruction. Knowing more Tier 2 words helps readers better understand the text they encounter, regardless of subject matter.

Researchers have compared a variety of approaches to Tier 2 word learning: when measuring expressive use of words and long-term retention, the consensus is that Tier 2 words require learning encounters where the word is presented in multiple contexts (Crist & Petrone, 1977;
Nash & Snowling, 2006; Stahl & Fairbanks, 2011). However, we know little about the ideal conditions for contextual word learning (CWL). The goal of Study 2 is to examine conditions that promote robust Tier 2 word learning from context. By robust, we mean the word meanings will be correctly inferred from the available context cues and the knowledge will be retained over time. We will focus on two dimensions of contextual word learning in this study: 1) contextual support (i.e., strength of cues to target word meaning) and 2) sequencing, or scaffolding, of contextual support (i.e., initially presenting easy contexts and progressively increasing the difficulty level).

6.1 Tier 2 Word Learning

Recent meta-analyses confirm the importance of ‘rich, extended practice’ (Beck, McKeown & Kucan, 2013), rather than incidental (non-directed) learning (Swanborn & de Glopper, 1999) or one-shot learning (‘fast mapping’; Munro, Baker, McGregor, Docking, & Arciuli, 2012). In this respect, learning of Tier 2 words is similar to learning in other domains: it typically requires multiple opportunities to learn and consolidate information into long-term memory (Jenkins, Stein, & Wysocki, 1984; Pavlik & Anderson, 2005). The optimal number of word learning trials is likely to vary for different learning and instructional tasks; however, previous studies have suggested that long-term retention could require 10-12 exposures during training (McKeown, Beck, Omanson, & Pople, 1985), though participants have shown significant retention over a week-long delay after 6 exposures (Frishkoff, Perfetti, Collins-Thompson, 2011) or even just 3 exposures (Lampinen & Faries, 1994, Frishkoff, Perfetti, Collins, Collins-Thompson, 2010). It is important to not only have multiple exposures, but also for those contexts to be diverse in order to reveal multiple shades of meaning. Repeated exposure to a word in a single context can help strengthen new word forms (Landi, Perfetti, Bolger, Dunlap, & Foorman, 2006; Adlof,
Frishkoff, Dandy, & Perfetti, 2015) and can reinforce meanings associated with a particular context. However, reading the same context multiple times is unlikely to reveal the nuanced meanings that are characteristic of Tier 2 words (Bolger, Balass, Landen, & Perfetti, 2008). The multifaceted and abstract nature of Tier 2 words makes learning and instruction more of a challenge: learners need multiple opportunities to read a word in context beyond a chance encounter in a grade-level text.

6.2 Contextual Support for Tier 2 Word Learning

Readers often encounter unfamiliar words while reading and can utilize the cues in the surrounding text, including words and grammatical markers or phrases, to infer the meaning of the unfamiliar word. This process is referred to as contextual word learning (CWL). The relative difficulty of lexical inferencing (guessing the meaning of the word) depends on multiple factors, including reader characteristics (e.g., age, reading skill level) and characteristics of the text itself. For example, length, readability and informational content of the text have all been shown to influence CWL outcomes (Baumann, Kame'enui, & Ash, 2003; Beck et al., 2013). In particular, word learning is more successful when texts are shorter and age- or skill-appropriate (Baumann, Kame'enui, & Ash, 2003) and when the target word is in close proximity to context cues, such as related words and phrases (e.g., Carnine, Kame'enui, & Coyle, 1984).

Beyond these general factors (which are controlled for in the present study), a key challenge is to weigh the costs and benefits associated with CWL through wide and independent reading and CWL through direct (classroom) instruction. In addition to the specific research aims for Study 2, a broader aim is to develop and test an approach to CWL that captures the best of both independent reading and direct instruction.
6.2.1 Arguments for independent reading.

Some authors have stressed the importance of independent reading for vocabulary acquisition. They note that the average child learns a staggering 3,000 words per year (±1,000; Nagy, Anderson & Herman, 1987), and it is impossible to learn all of these words through direct instruction (e.g., Scott, Nagy, & Flinspach, 2008). Through their own reading, readers are exposed to a wide variety of words in a diverse set of contexts. As previously noted, exposure to diverse contexts is essential for acquiring Tier 2 word knowledge. However, evidence concerning the efficacy of CWL through independent reading has been mixed.

A meta-analysis conducted by Kuhn and Stahl (1998) showed that children who practiced contextual word learning, even without guidance or instruction, improved at deriving words from context: suggesting that perhaps the practice independent reading provides is sufficient for CWL. However, researchers have noted several potential problems with vocabulary acquisition through independent reading. First, authentic contexts (i.e., contexts that are not experimentally designed, for example, literature or classroom textbooks) often provide few, if any, cues to the meaning of an unknown word (Swanborn & de Glopper, 1999; Schatz & Baldwin, 1986). When authentic texts do provide cues, the cues are often inconsistent or outright misleading (Beck, McKeown, & McCaslin, 1983; Frishkoff et al., 2008; Schatz & Baldwin, 1986). Second, CWL involves more than strengthening of simple associations: it involves skills, such as lexical inferencing, which vary across readers. Some research suggests that less-skilled readers experience smaller gains from non-directed reading, as compared with skilled readers (Cain & Oakhill, 2011; Cain, Oakhill, & Lemmon, 2004; Cain, Oakhill, & Elbro, 2003; Liu & Nation, 1985).
6.2.2 Arguments for direct instruction.

An alternative — or supplement — to independent reading is direct instruction. For example, Elements of Reading®: Vocabulary (EOR:V) is a structured curriculum that teaches a small number of words each week through intensive class exercises (Beck & McKeown, 2004). During group discussions of assigned readings, the teacher provides student-friendly definitions where needed, and leads short (~5-15 mins) exercises that illustrate correct and incorrect uses of the word and show how the word is (or is not) related to familiar words and concepts. Direct instruction has several advantages over independent reading: it gives unambiguous support for learning a word, and multiple opportunities to practice retrieving the word from memory. Two independent studies using quasi-experimental randomized trials found that EOR:V leads to gains on global measures of vocabulary and comprehension, compared with controls (Apthorp, 2005; Resendez, Sridiharan, & Azin, 2006; but see Apthorp et al., 2012, for more equivocal findings). However, there are also inherent limitations in direct instruction. Most importantly, while it can be effective for accurate word knowledge, direct classroom instruction can only teach a fraction of the words that need to be learned (Sternberg, 1987). In addition, it requires teacher time to create new exercises, and class time to implement.

To summarize, independent reading and direct instruction may be complementary approaches to word learning. Independent reading can take place outside the classroom; thus, it is potentially more effective over the long-term. A downside to independent reading is that authentic contexts do not provide strong or consistent cues to meaning, and may lead to incorrect or incomplete word knowledge. Direct instruction leads to accurate word knowledge, but only for a
small fraction of words. Considering these tradeoffs, it seems clear that independent reading and direct instruction are both necessary to support robust learning of Tier 2 words (Kamil et al., 2008; Slavin, Lake, Chambers, Cheung, & Davis, 2010).

6.2.3 An intelligent tutor for optimal CWL.

A third approach to CWL uses computer-aided instruction to address the trade-offs between independent reading and direct classroom instruction, and to leverage the advantages of each in order to promote optimal word learning outcomes. For example, Frishkoff, Collins-Thompson, Hodges, and Crossley (rev. under review) describe an intelligent tutoring system (ITS) that provides high-quality contexts for CWL and structures presentation of contexts over time to optimize learning and retention. Compared with independent learning, this approach may be more effective because contexts are selected to provide high-quality information about the target words. In this way, the advantage of varied exposures is retained, while the disadvantage of uninformative contexts is reduced or even eliminated. Compared with direct instruction, an ITS may be more efficient because it can focus on words that are unfamiliar to each individual, which is difficult to accomplish with group instruction. The ITS can also support CWL outside of class: the student can use the ITS as part of an independent learning activity. It also supports constant active engagement with the CWL process by requiring the learner to provide responses and answer questions about their knowledge of a word as it develops over time. Neither independent reading nor direct instruction are able to provide this same level of active processing for every student or for every targeted word, though research shows that testing (especially recall tasks) support long-term learning (e.g., Roediger & Karpicke, 2006).

Another advantage of the ITS is that it can serve as a research platform, as well as an educational tool. Our understanding of what conditions lead to optimal CWL, and why, is still in its
infancy. Since the ITS allows for precise control over stimulus materials and presentation conditions (e.g., timing, sequencing), it can support authentic classroom research. This means we can test specific cognitive theories of CWL while simultaneously testing the practical value of the ITS.

6.3 Scaffolding of Support in CWL

Contextual word learning relies on lexical inferencing, where the reader guesses the meaning of an unknown word from the surrounding cues (Dulin, 1970). These cues include phrase and clause-level markers that indicate how two words are related. In the sentence “He had spinach, not arugula,” for example, the word ‘not’ implies that spinach and arugula are things that can be compared; hence, they are probably members of the same semantic category. The meaning of a known (source) word also provides cues to the meaning of a novel (target) word. In this case, knowing the meaning of ‘spinach’ and learning that arugula and spinach are members of the same category supports the inference that arugula is a leafy green vegetable (Frishkoff, Collins-Thompson, Hodges, & Crossley, rev. under review). Over the past 50 years, researchers have identified a variety of context cue types, including cues to synonymy, cause-and-effect relationships between two words, and contrastive cues, as illustrated in the above example (Dulin, 1970).

When readability and length of contexts are controlled, a major predictor of CWL outcomes is context constraint. Context constraint is operationally defined by the cloze procedure (Taylor, 1953). People are presented with a partial context, where a word has been removed and replaced with a blank. The task for each respondent is to provide the missing word. The constraint of a context reflects how strongly it elicits a particular word or set of words (Taylor, 1953). High-constraint contexts elicit a small number of completions, suggesting that they prime
a small set of word representations in memory (e.g., She asked him for an answer, but he kept _____ and refused to speak.). If there is an unknown word in a High-constraint context, it is easy to guess what the word means: it is likely to be a near-synonym for the modal (most common) response on the cloze task. Low-constraint contexts elicit more varied completions, consistent with weaker or more diffuse activation of semantic memory. Thus, they provide minimal support for lexical inferencing (egg., Some people are more _____ than others.). High-constraint contexts provide strong support for learning, because they limit the set of plausible meanings that can be associated with the unknown word. Low-constraint contexts provide very little support: they are semantically underspecified, meaning there are few limits on what the unknown word could mean.

However, context constraint is not always high or low; rather, it varies from high to medium to low. This variation raises some important questions about core cognitive processes in CWL. For example: Should contexts always make it easy to guess a word's meaning? Are there conditions where it is actually better to provide more challenging (i.e., less informative) contexts? If we take a page from the cognitive literature on learning and memory, the answer should be yes.

Schmidt and Bjork (1992) proposed a ‘Desirable Difficulties’ framework, which explains the benefits of presenting cognitively challenging problems during learning and testing. This framework has been used to account for a wide range of effects in the cognitive science of learning and memory, including interleaving and spacing of practice, testing-for-learning, and the generation effect (Schmidt & Bjork, 1992). For all of these effects, better outcomes result from the more challenging conditions; for example, students completing a recall test rather than simply re-reading study materials before an exam perform better on the exam. Of course, the chal-
lenge must be appropriate to the student's level: a problem that is too hard can lead to short-term failure or long-term avoidance of the problem (Yerkes & Dodson, 1908; Vygotsky, 1962). Recently, the problem of determining the optimal degree of difficulty for a particular task has been termed the ‘Assistance Dilemma’ (Koedinger & Aleven, 2007). According to Koedinger and associates, the problem is finding the balance between how much information to provide and how much information to withhold (Koedinger, Booth, & Klahr, 2013). This finding has been replicated extensively in associative word learning (e.g., learning to associate a word with a picture or definition, Ebbinghaus, 1885, for a review, see Delaney, Verkoeijen, Spirgel, 2010 or Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006), though not in contextual word learning.

There are reasons to think that the Desirable Difficulties principle, or the Assistance Dilemma, may extend to CWL. For example, contextual word learning studies using eye tracking measures have shown that learners spend less time looking at novel (target) words in highly constrained contexts as compared with less constrained contexts (Rayner & Well, 1996), consistent with the idea that high constraint contexts are less engaging. For robust contextual word learning, perhaps more challenging contexts (medium or low constraint) lead to better outcomes. Prior work has found that high-constraint contexts lead to more successful word learning than low-constraint contexts (Beck, McKeown, & McCaslin, 1983; Chaffin, 1997; Chaffin, Morris, & Seely, 2001; Frishkoff, Perfetti, & Collins-Thompson, 2010, 2011). However, high-constraint contexts do not lead to optimal retention of new word knowledge, where the knowledge is retained over a significant period of time (i.e., one week or more, e.g., Frishkoff, Perfetti, Collins-Thompson, 2011).

On the other hand, Lampinen and Faries (1994) found that students who were exposed to a new word in three medium-constraint contexts outperformed students who saw these same
words in three high-constraint contexts on both a sentence verification task and a definition generation task when tested immediately after training. Furthermore, the difference remained after a 1-week delay. This result is consistent with the Desirable difficulties framework: when a training example provides minimal or incomplete cues, the learner must engage in deeper, more effortful processing (e.g., inferencing, elaboration, or retrieval of information from long-term memory).

6.3.1 Scaffolding of contextual support.

As discussed previously, it is necessary to present Tier 2 words in multiple contexts to reveal the nuanced meaning of the word and cement it into long-term memory (Bolger et al., 2008, Jenkins et al., 1984). Accepting that requirement, an important next question is how best to sequence the presentation of these contexts; what is as an ideal context on trial 1 may not be ideal on trial 4. This dynamic component is the motivation for scaffolding of support (i.e., providing strong support on early trials and decreasing support over time). Previous studies have found that scaffolding supports optimal learning across domains, including math and science, as well as vocabulary and reading comprehension (Rittle-Johnson & Koedinger, 2005; Graesser, McNamara, & Van Lehn, 2005). Scaffolding may be effective because high levels of support on early trials minimize error (i.e., incorrect coding of new knowledge), while subsequent fading of support gives rise to desirable difficulties (Schmidt & Bjork, 1992).

To our knowledge, Lampinen and Faries (1994) is the only study that has examined scaffolding of support in CWL. In addition to an all-high constraint and an all-medium constraint condition, their study included a third mixed constraint condition, in which subjects were exposed to one high-constraint context, followed by one moderately constraining context, and one low-constraint context (a scaffolded sequence of varying contextual constraint). On the immediate post-test, students in the mixed condition outperformed students in the all-high constraint
condition on both a sentence verification task and a definition generation task. A similar pattern was observed on a 1-week delayed post-test. The authors suggest that scaffolding of support is helpful because it promotes more effortful processing as compared with the all-high condition. Interestingly, outcomes were not different between the scaffolded and all-medium conditions. The comparable outcomes of the all-medium and scaffolded condition led the authors to conclude that it was the presence of medium-constraint contexts that explained the advantages of the scaffolded condition over the all-high condition, not the sequencing of contexts from high to medium to low.

6.4 Overview of Study 2

The goal of the present study is to examine conditions that promote robust CWL, and in particular, to replicate and extend the findings of Lampinen and Faries (1994) - though we are using different measures of word knowledge. The overarching question is: Does a particular sequencing of high-, medium-, and low-constraint contexts lead to more robust learning from context than other sequences? To address this question, we assigned each novel target word to one of four context training conditions: (1) AllHigh, which consisted of four maximally supportive (i.e., high-constraint) contexts; (2) AllMedium, which consisted of four moderately supportive (i.e., medium-constraint) contexts; (3) Scaffolded, which provided one high-constraint context, followed by successively less constraining contexts; or (4) Ascending, which was the inverse of the Scaffolded condition (i.e., the first context was minimally informative, and subsequent contexts provided increasing support). We assessed knowledge of the target words at three separate time points: a pretest, an immediate posttest following the training, and a delayed posttest one week after the training.
We had two main hypotheses. First, we predicted that a series of highly supportive (All-High) contexts would lead to better immediate learning, but worse retention, compared with a series of moderately supportive (AllMedium) contexts based on previous work on contextual word learning (Beck, McKeown, & McCaslin, 1983; Frishkoff, Perfetti, & Collins–Thompson, 2010, 2011; we refer to this as the Context Constraint Hypothesis). We hypothesized that the definition generation task, which is performed throughout the contextual word learning training, engages two processes: lexical inferencing (i.e., using context cues to infer meaning, particularly on early trials) and memory retrieval (i.e., reactivating word representations that were learned on previous trials). As discussed earlier, the AllHigh condition could reduce the need for memory retrieval on later trials: if the participant can readily guess the word's meaning from the context, retrieval is unnecessary. Given the importance of retrieval in long-term consolidation of memories (e.g., Delaney, Verkoeijin, Spirgel, 2010), this could lead to sharp decreases in performance from the immediate to the delayed posttest, as compared with the AllMedium condition. On the other hand, high-constraint contexts could give an advantage for immediate gains in word knowledge: high-constraint contexts consistently limit the likely meaning of a word, whereas moderately constraining contexts allow for more varied meanings to be inferred across contexts.

Our second prediction was that scaffolding of contextual support would lead to both optimal learning and retention (we refer to this as the Scaffolding Hypothesis). This prediction has two parts. First, we predicted that scaffolding would lead to smaller drops in performance from the immediate to the delayed post-test as compared with the AllHigh condition, as in Lampinen and Faries (1994). However, in contrast with their findings, we also expected to find an advantage of scaffolding for initial learning (from pre- to immediate post-test) as seen in other learning domains (Rittle-Johnson & Koedinger, 2005; Graesser, McNamara, & Van Lehn, 2005).
The presence of an initial high-constraint context should reduce the likelihood of incorrect inferences, which should translate to better performance for the AllHigh and Scaffolded conditions, as compared with the AllMedium condition, on the immediate post-test.

In addition, we included a fourth condition, which includes the same mixture of high-, medium-, and low-constraint contexts as the Scaffolded condition, but reverses the order (Ascending). Differences in performance for the Scaffolded versus Ascending conditions would provide evidence that it is not simply the presence of medium-constraint contexts -- or even the mixture of high, medium and low constraint contexts -- that matters for robust CWL as suggested by Lampinen & Faries (1994). This would provide strong evidence for our Scaffolding Hypothesis: that it is the particular sequencing of contexts from high to low support that leads to optimal long-term outcomes.

7 STUDY 2 METHOD

7.1 Participants

One hundred English-speaking adults (77 females; $M = 21$ years, $SD = 3.5$ years) were recruited from the human subjects pool at Georgia State University and completed at least a portion of the pretest session. A subset of these participants $(n = 45)$ completed all three sessions (pretest, training and immediate post-test, and delayed post-test). Three participants were excluded from the final analysis: two of these three participants completed the delayed post-test session outside the time window (1 week, $\pm 1$ day), and one participant received extra practice due to experimenter error. After exclusion of these participants, there were 42 datasets remaining ($n = 36$ female), all of which were used in the final analysis.

All participants completed the Nelson-Denny Reading and Vocabulary test at the pretest session. This is a standardized measure that provides three scores: vocabulary, reading compre-
hension, and a total score (Form G, Brown, Fishco & Hanna, 1993). The timed 15-minute vocabulary section contains 80 vocabulary items with five answer choices each. The timed 20-minute comprehension section contains seven reading passages and 38 questions with five answer choices each. Participants included in our analyses had a mean vocabulary standard score of 213.88 (SD = 29.02), and a mean comprehension standard score of 213.86 (SD = 26.77). Importantly, there was no difference when comparing the Nelson Denny total standard score (vocabulary standard score + reading comprehension standard score) between the participants who completed the third session and those who did not, $t(96) = -.18$, $p = .86$ (note that not all of the original 100 participants sufficiently completed the Nelson Denny for inclusion in this analysis).

7.2 Experiment Stimuli

For this study, we created a set of target words and a corpus of sentence-long contexts for the contextual word learning training.

7.2.1 Target words.

The target words include 48 very rare English words, appearing in written contexts less than once per 1 million written words (Wilson, 1988). Most of the very rare words were used in previous studies of CWL (Frishkoff, Collins-Thompson, Perfetti, & Callan, 2008; Frishkoff, Perfetti, Collins-Thompson, 2011) and results have repeatedly shown that college-educated native English speaking adults are no better than chance in selecting the correct meanings of these words (Frishkoff, Collins-Thompson, Perfetti, & Callan, 2008), that they regard these words as very unfamiliar (close to '1' on a scale of '1' to '3'), and that they fail to distinguish very rare words from pseudowords on a lexical decision task (Frishkoff, Perfetti, & Westbury, 2009). Thus, we are confident that participants will have no prior knowledge of target words for learning in our task. The advantage of using very rare words, rather than non-words, is that variability
in baseline knowledge is relatively small across participants (and close to zero). At the same time, these particular rare words are abstract, multidimensional, and polysemous, making them comparable to Tier 2 words. In addition to the rare words, 30 low-frequency, familiar words were included on the assessments (but not the training), so that participants would not become overly frustrated (especially on the pre-test).

Table 2. Properties of target and non-target words

<table>
<thead>
<tr>
<th>Word Group</th>
<th>Number of words</th>
<th>Length in letters</th>
<th>Text frequency per million written words</th>
<th>Spoken frequency per ~190,000 words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Familiar</td>
<td>30</td>
<td>6.9 (1.3)</td>
<td>6.7 (8.0)</td>
<td>0.3 (0.5)</td>
</tr>
<tr>
<td>Rare Untrained</td>
<td>15</td>
<td>7.6 (1.2)*</td>
<td>0.1 (0.0)</td>
<td>0.0 (0.0)</td>
</tr>
<tr>
<td>Rare Trained</td>
<td>48</td>
<td>6.8 (1.1)</td>
<td>0.0 (1.0)</td>
<td>0.0 (0.0)</td>
</tr>
</tbody>
</table>

*Note. (Standard deviation in parentheses). Significant difference between the length of the trained and untrained rare words, \( p = .01 \). Frequency of occurrence in written language derived from Francis and Kučera (1967) norms, ranging from 0-69971 occurrences per million words. Frequency of occurrence in spoken language derived from the London-Lund Corpus of English Conversation by Brown (1984) containing 191,918 spoken words, word frequency ranges from 0 - 6833, \( M = 35, SD = of 252 \). All values returned using the MRC psycholinguistic database (Wilson, 1988).

7.2.2 Training contexts.

For each target word, there were four low constraint sentences, four medium constraint sentences, and four high constraint sentences, which were selected from a larger corpus of sentences that we created and normed. The high, medium, and low sentences were equated for sentence length, syntactic complexity, and readability (approximately 4th-grade level) as well as the position of the target word within the sentence. Classification of each sentence as high, medium,
or low constraint was based on data from a norming ‘cloze’ experiment. In a cloze task, target words are removed, and participants are asked to fill in the blank with the word they felt best completes the sentence (Taylor, 1953). A separate group of participants (N = 47, 32 female, \( M_{\text{age}} = 19.7 \) years) completed this cloze sentence norming task, yielding an average of 23 responses per sentence. For example, the participant would see “Ginger's pencil was so \___________\ she accidentally stabbed her finger with it.” Measures of sentential constraint were computed, based on the number and density of unique words generated on the cloze task for each sentence. These measures were then entered into K-means clustering, specifying three (for high, medium and low) as the desired number of clusters. High constraint sentences elicited between 1 and 7 distinct cloze completions, for example, sharp and sharpened would not be considered distinct, but sharp and pointy would be considered distinct responses (average distinct responses = 4). Medium-constraint sentences elicited between 7 and 12 unique responses (average = 9). Low-constraint sentences elicited between 12 and 26 distinct responses (average = 15).

**Table 3. Context constraint examples**

<table>
<thead>
<tr>
<th>Target Word</th>
<th>Low Constraint</th>
<th>Medium Constraint</th>
<th>High Constraint</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Conticent”</td>
<td>Some people are more <em>conticent</em> than others.</td>
<td>The sign said you had to be <em>conticent</em> as you walked around the castle.</td>
<td>She asked him for an answer, but he kept <em>conticent</em> and refused to speak.</td>
</tr>
<tr>
<td>“Versute”</td>
<td>I know some <em>versute</em> people, but I am not one of them.</td>
<td>Jessica could not think of any <em>versute</em> comments, so she changed the subject.</td>
<td>We need someone who is <em>versute</em> to find a solution.</td>
</tr>
</tbody>
</table>

### 7.3 Vocabulary assessments

Two assessments of vocabulary knowledge were administered that will be analyzed within this study. The same words were in both the familiarity and the synonym tasks: non-target
words – including low frequency words to reduce participants’ frustration (particularly at the pretest) and untrained rare words, and target words - the trained rare words (refer to Table 2 for word group characteristics).

7.3.1 Familiarity task.

In this task, the participant was presented with each word one at a time and asked to rate their familiarity with each word as '1' (I don't remember having seen this word before), '2' (I have seen this word before but I don't know what it means) or '3' (I have seen this word before and I think I know what it means). This task has been validated in prior word learning studies (Wesche, 1996). Figure 4 shows the mean percent of familiarity ratings for each of the word types (familiar, rare trained, and rare untrained) at pretest and the delayed posttest. Participants rated most (~%80) of the rare words as a '1' at the pretest, but at the posttest most (80-90%) of the rare words were rated as either '2' or '3'. Since the untrained words were seen at every administration of the synonym task as well as on the familiarity task, it is unsurprising that the majority of the untrained words were rated as '2'. It is important to note that participants rated more of the trained than untrained rare words as '3', indicating an effect of context word learning training on their awareness of their own word knowledge. The ratings for the familiar words were relatively similar between the pretest and the delayed posttest. These ratings reflect metacognitive awareness of the participants’ word knowledge for each word group.
7.3.2 Synonym task.

In this task, the participant saw each target word with five other words (same part-of-speech) appearing below the target. Participants were instructed to choose the word closest in meaning to the target word. The answer choices followed a pattern: the correct answer, an antonym of the correct answer, another synonym/antonym pair, and another word of similar length and frequency to the rest of the options. For example, for the target word ‘continent,’ the choices included ‘silent’, ‘talkative’, ‘mournful’, ‘happy’, and selective, with ‘silent’ as the correct answer choice. The Synonym Task included all of the words – the 48 trained target rare words, 15 untrained rare words, and 30 familiar low-frequency words (see Table 2).

In addition to the Familiarity Task and the Synonym Task, participants completed two other tasks: a definition generation task to ensure active participation throughout the contextual word learning phase, and a priming task. In the definition generation task, the participant saw each target word presented in isolation, and was asked to provide a one-word definition. In the
priming task, two words were presented consecutively and the participant responded yes if they thought the words were related and no if they were unrelated. The definition generation task will be analyzed at a later date though the priming task will not be analyzed due to technical difficulties during presentation.

7.4 Data Collection Procedure

Participants attended three separate sessions on three separate days. In session 1 they completed the Nelson Denny reading and vocabulary tests, and several pretest assessments used to establish baseline scores. During the second session, participants completed the CWL training task, followed by an immediate post-test. In the training task, target words were randomly assigned to condition for each participant, such that there were 12 target words assigned to each of the four within-subjects conditions. Each word was presented in four different contexts, for a total of 192 trials (4 conditions X 12 words per condition X 4 contexts per word = 192 trials). The four experimental conditions were as follows: all high constraint (AllHigh), all medium constraint (AllMedium), decreasing contextual constraint (Scaffolded), and increasing contextual constraint (Ascending). There were also 15 untrained rare words that were presented on the Familiarity Task and the Synonym task that were never presented in the context training (UntrainedRare). Session 3, the final session, assessed final retention of the targeted vocabulary terms one week after session 2.

The sessions were conducted at Georgia State University in a quiet room. Only trained research assistants following a script administered the testing sessions. For session 3, some participants were permitted to log onto the online platform from home in an effort to retain as many participants as possible.
Figure 5. Overview of contextual word learning study.
This is the design of the study, with a pretest, immediate posttest and delayed posttest administered on three separate days. The context training was presented during session 2, and then immediately followed by the immediate posttest. While the time between Sessions 1 and 2 was not strictly controlled (ranging from 1 to 7 days), the time between Sessions 2 and 3 was scheduled at 1 week (±1 day).

7.4.1 Session 1: Pretest.

During this initial session, participants were informed of the purpose of the study: to understand contextual word learning. They were told that they would be completing similar tasks measuring their word knowledge over three separate sessions. They completed the consent forms, and the Nelson-Denny vocabulary and reading comprehension assessment. Then, participants were assigned a unique ID number and instructed to log onto the online intelligent tutoring system (ITS). Three tasks were administered, assessing different aspects of the participants’ knowledge of the target words.

7.4.2 Session 2: Training + Immediate posttest.

In the second session, participants completed the context training task and the immediate post-tests. The timing between the initial session and the second session was not strictly controlled. During the training task, participants were presented with one sentence at a time containing a target word. After reading each sentence, the participant provided a one-word definition for the target word. There were 192 trials, such that each of the 48 words was presented a total of 4
times (4 training conditions X 12 words per training condition X 4 contexts per word = 192 trials). It is important to note that the four sentences per word were not presented consecutively; the spacing between the four sentences was consistent with approximately 25-30 sentences between exposures to a particular target word.

There were four constraint conditions as a within subjects variable. This means that for each participant, 12 words were assigned to each of the four conditions (with the assignment of words to condition counterbalanced across subjects). The conditions are as follows: (1) all high constraint (AllHigh), which included four different high constraint sentences; (2) all medium constraint (AllMedium), which included four different medium constraint sentences; (3) descending constraint (Scaffolded), which included one high constraint, followed by two medium constraint, and a low constraint sentence; and (4) ascending constraint (Ascending), which was the opposite of the Scaffolded condition with one low, followed by two medium and then a high constraint sentence. At no point in time did the participant receive feedback about their responses. After completing all 192 training trials, participants completed the immediate post-test. This posttest consisted of the synonym judgment task and the definition generation task. These tasks are identical to the baseline tasks in the first session.

7.4.3 Session 3: Delayed posttest.

Participants returned for the final session one week (±1 day) after completing the second session. During this session, they completed the delayed post-test to measure final retention of the target words. The post-test tasks included the synonym judgment task, the familiarity judgment task, and the definition generation task. The priming task was also administered in this session.
7.5 Procedure for Data Analysis

For the purposes of this study, only participants who completed all three sessions were included for analysis. Scores on the Synonym Task were analyzed using a 3 (Session) by 5 (Condition) repeated-measures analysis of variance (ANOVA). Session (3: pretest, immediate posttest, and delayed posttest) and Condition (4: AllHigh, AllMedium, Scaffolded, and Ascending, Untrained) were both within-subjects variables. In order to examine the interaction of Session and Condition, we conducted a simple main effects analysis.

Since there were five answer options for each trial on the Synonym Task, an average proportion of correct responses at .20 represents chance. We expected accuracy on the immediate post-test to be highest for the AllHigh condition. However, on the delayed post-test, we expected accuracy to be highest on words trained in the AllMedium or the Scaffolded condition compared to the AllHigh and Ascending conditions, though we did not necessarily expect a significant difference between AllMedium and the Scaffolded condition.

8 STUDY 2 RESULTS

The synonym task was the primary outcome measure for Study 2. Table 4 below shows the means on the synonym task for all three of the words groups, familiar, rare untrained, and rare trained word. The analyses did not include the familiar words, only the rare trained and rare untrained words.
Table 4. Mean proportion correct on the Synonym Task

<table>
<thead>
<tr>
<th></th>
<th>Pretest Mean(SD)</th>
<th>Immediate Posttest Mean(SD)</th>
<th>Delayed Posttest Mean(SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Familiar Untrained</td>
<td>.83(.38)</td>
<td>.84(.36)</td>
<td>.84(.37)</td>
</tr>
<tr>
<td>Rare Untrained</td>
<td>.23(.42)</td>
<td>.24(.43)</td>
<td>.28(.45)</td>
</tr>
<tr>
<td>Rare Trained (ALL)</td>
<td>.23(.42)</td>
<td>.56(.50)</td>
<td>.47(.50)</td>
</tr>
</tbody>
</table>

Note. (Standard deviation in parentheses).

Scores on the Synonym task were analyzed using a 3 (Session) by 5 (Condition) repeated-measures analysis of variance (ANOVA). Session (3: pretest, immediate posttest, and delayed posttest) and Condition (5: AllHigh, AllMedium, Scaffolded, Ascending, and Untrained) were both within-subjects variables. We found a main effect of Session, \(F(2, 82) = 122.93, p < .001, \eta^2 = .83\). Figure 6 shows that while scores were close to chance (.20) at Session 1, accuracy increased for the trained rare words and then decreased somewhat at Session 3. There was also a main effect of Condition, \(F(4, 164) = 27.08, p < .001, \eta^2 = .40\). More importantly, there was a significant interaction of Condition by Session, \(F(8, 328) = 15.39, p < .001, \eta^p^2 = .27\).
Figure 6. Synonym task scores
Mean proportion correct of 0.20 represents chance performance since there were five answer choices.

Since the interaction of session and condition was significant, we did a simple main effects analysis. At the immediate posttest, the AllHigh condition did not lead to significantly better accuracy than the AllMedium condition, \([F(1,41) = 1.04, p = .31]\), or the Scaffolded condition, \([F(1,41) = .43, p = .52]\).

At the delayed posttest, there was a significant difference between the AllHigh condition as compared with the Scaffolded condition, such that the Scaffolded condition resulted in higher accuracy, \([F(1,41) = 10.13, p = .003]\). There was a trend towards a significant difference between the AllHigh and the AllMedium condition \([F(1,41) = 3.34, p = .08]\), such that the AllMedium condition resulted in slightly higher accuracy. There was no difference between the AllMedium condition as compared with the Scaffolded condition \([F(1,41) = 2.06, p = .16]\), but
there was a slight trend towards an advantage for the Scaffolded as compared with the Ascending condition, \( F(1,41) = 2.83, p = .10 \).

Figure 7 displays difference scores for the experimental conditions as another way to visualize the results: light-colored bars indicate learning (i.e., changes from pretest to immediate posttest); dark-colored bars indicate forgetting (i.e., changes from immediate to delayed posttest). This illustrates the comparable learning between the AllHigh and Scaffolded conditions, and the comparable forgetting between the AllMedium and the Scaffolded conditions.

**Figure 7.** Difference scores as measures of learning and forgetting (Error bars indicate standard error).
9 STUDY 2 DISCUSSION

Study 2 examined learning and retention of rare (Tier 2 like) words as a function of context constraint and sequencing of high-, medium-, and low-constraint sentences. We predicted that a series of medium-constraint sentences (AllMedium condition) would lead to smaller gains but superior retention compared to a series of maximally supportive contexts (AllHigh condition; Context Constraint Hypothesis). We also predicted that scaffolding of contextual support would produce strong gains on the immediate posttest, equal to or better than the AllHigh condition, and that these gains would be retained at least as well as the AllMedium condition (Scaffolding Hypothesis). Our results provided support for both of our hypotheses.

Our results are similar to the results from Lampinen and Faries (1994), who showed that new words are better retained when they are presented in medium- as opposed to high-constraint contexts. Though the difference between these two conditions did not reach significance in this study, the trend was the same. We also did not find the expected advantage for the AllHigh condition at the immediate posttest. Since Lampinen and Faries used different outcome measures (evaluation of correct usage of target words in a sentence, providing a short definition for each word) the similar findings indicate the effects are consistent regardless of the assessment task demands. The advantage for medium-constraint contexts may be that they provide enough information to support accurate inferences, but not so much that the answer is immediately obvious. This idea is consistent with eye tracking studies, which have shown that learners fixate less and sometimes even skip, words that are highly predictable from context (Rayner & Well, 1996). While this behavior makes sense when a word is already familiar, it could be detrimental for CWL. If the third or fourth exposure to a new word involves a high-constraint context, then meaning of the word is obvious from the context and there is no need to recall knowledge from
previous encounters. Since memory retrieval is important for long-term consolidation, this could be viewed as a missed opportunity (Pashler et al., 2007). Also similar to Lampinen and Faries (1994), the present study showed that scaffolding of support can be effectively implemented in CWL by ordering contexts from high to medium to low constraint.

There are some differences in study methodology and outcomes. Lampinen and Faries (1994) found no differences between their AllMedium and Scaffolded conditions, and nor did we. However, they found that on both the immediate and delayed post-tests, students in the AllMedium and Scaffolded (what they termed the mixed condition) conditions outperformed students in the AllHigh condition. We did not find the same advantage for AllMedium and Scaffolded at both time points; only at the delayed posttest did the Scaffolded condition result in significantly greater accuracy than the AllHigh condition, with a slight (non-significant) advantage for the AllMedium as compared with the AllHigh. This difference between the previous study and the current work could be explained in a number of ways. First, it is possible that variables such as sentence length or readability differed across conditions in Lampinen and Faries (1994), while we equated those variables across our context constraint groups. It is also possible that their medium-constraint sentences were less variable or even more informative than the medium-constraint sentences used in the present study. Finally, it is important to note that condition was a between-subjects variable in their study, while it was a within-subjects condition in this study. Perhaps learners who were assigned to the AllMedium condition used different strategies than learners in the other conditions: because all of the contexts were somewhat challenging, they may have worked harder to compare and contrast cues to meaning across trials.

The present study also included a fourth, Ascending, condition, for comparison with the Scaffolded condition. Given that the Scaffolded condition resulted in a trend towards a greater
proportion of correct answers on the Synonym Task, we can tentatively conclude it is the scaffolded presentation sequence, not just variability of constraint, that matters for robust CWL. This finding does not support Lampinen and Faries’ conclusion that the advantage of the Scaffolded condition was solely the inclusion of the medium-constraint sentences. If this were true, then the sequencing would have no effect since both the Scaffolded and Ascending conditions included two medium-constraint sentences. Since the effects are small, it is unclear if there is a true difference between the two conditions.

While this study did provide evidence supporting the scaffolding hypothesis, but not the context constraint hypothesis, there were some limitations to this work. First of all, there were no untrained rare words on the delayed posttest that were not previously seen. All of the untrained rare words were included on all of the non-training tasks (though the familiarity task was only administered at the pretest and the delayed posttest). This means even though those words were not seen in context, they were still somewhat familiar to the participants by the delayed posttest because the word form had been viewed multiple times. This study also did not contain a random control condition. This was by design, since we did not want to have cases in a random condition where the first encounter was a high-constraint context: this would too closely mimic the Scaffolded condition. Instead, we included the Ascending condition as a comparison for the Scaffolded condition – while the Ascending condition contained the same context levels as the Scaffolded condition, the presentation was from low to high constraint rather than high to low. A future study may include a pseudo-random condition, where the first exposure is a low-constraint context, but the subsequent exposures are randomized. Also, due to high attrition rates, the study
was slightly underpowered, meaning that not all effects could reliably be detected. To address this issue, further data are being collected so that analyses can be rerun with greater confidence in the results.

Beyond addressing the shortcomings of this study, we are also interested in answering further questions about contextual word learning. One way to do this is to examine individual differences in reading comprehension and vocabulary skill and the effect of training conditions on word learning from context. Children of varying skill levels are able to successfully learn words from context (Van Daalen-Kapteijns, Elshout-Mohr, & de Glopper, 2001; Nash & Snowling, 2006). However, previous work has indicated that some manipulations, such as the spacing between encounters, may have differential effects depending on skill level. Frishkoff et al. (2008) presented high- and low-skilled readers with Tier 2 target words (e.g., 'salubrious') in six contexts that either reinforced the correct meaning of the target or misused the word in a way consistent with the meaning of a similar-sounding low-frequency word (e.g., 'salacious'). When pre-test scores on a semantic judgment task were compared with scores on a delayed post-test, there was a marginal effect of spacing ($p < .06$). However, there was an interaction with reading skill: high-skilled readers showed a robust spacing effect ($p < .01$), where there was an advantage for encounters being more widely spaced, whereas low-skilled readers did not ($p > .7$). In sum, this study showed that while the manipulation of wider spacing provided an advantage to one group, the high skill readers, there was no advantage for the low skill readers (though there was also not a detrimental effect either). Future work can expand on this spacing factor by providing multiple study session and manipulating the time between the study sessions.

Pashler and associates have examined how learning can be distributed over multiple study sessions in order to optimize retention on a delayed post-test (e.g., Cepeda et al., 2009).
Their work suggests a non-monotonic relationship between the **interstudy interval (ISI)** — that is, the time between study sessions 1 and 2 — and the **study–test retention interval (RI)**, i.e., the time that separates the final study session and the final (delayed) post-test. When the RI is short, accuracy is higher when there is less time between study sessions, but when RI is long (weeks, months, years), it is better to have more time between study sessions, that is, a longer ISI (Glenberg & Lehmann, 1980; Bahrick & Phelps, 1987). A recent review (Cepeda et al., 2006), indicates the optimal ISI equals ~10-20% of the RI. In addition, while most studies have focused on associative learning, similar effects have been observed for acquisition of grammatical or mathematical rules, which — like CWL — engage inferential processes that go beyond associative learning (Carpenter et al., 2012). Future work can test we can test if this holds true for contextual word learning and if there are individual differences based on reading skill level.

### 10 GENERAL DISCUSSION

There are thousands of words children need to learn to be successful in school (Hirsch, 2003). It is vital for children to have a strong command of vocabulary before entering school and embarking on the challenge of learning to read written words (Chall, 1983). These two studies serve as first steps in understanding how to support children’s word learning at both the pre- and post-literate stages.

For preliterate children, gesture, particularly iconic gestures, can provide a support for learning increasingly abstract words. In order for this to be effectively utilized, we must know the age at which children comprehend iconic gestures and if comprehension emerges earlier for some subtypes of iconic gesture in order to provide the most informative gestures to children that can aid in word learning. In Study 1, we found that by age 2, children comprehend iconic gestures, but only iconic gestures that convey action information. By age 3, children comprehend
iconic gestures that convey action or attribute information about the referent. This suggests that different types of iconicity might be crucial in helping with word learning at the early ages.

For literate children, written text provides an opportunity to learn words. In Study 2, we focused on identifying the conditions that support word learning from context. In Study 2, we addressed competing hypotheses of the ideal level of context constraint for robust (long-term) Tier 2 word learning. We focused specifically on Tier 2 words since these words occur in texts covering a variety of subjects, and thus can support comprehension of a wider variety of texts, as opposed to more subject-specific vocabulary. We found that for learning gains, a condition with a series of all highly informative contexts was better than a series of all moderately informative contexts. However, knowledge loss was also greater in the AllHigh condition compared to the AllMedium condition. The Scaffolded condition resulted in the best of both worlds; learning was just as good as in the AllHigh condition and forgetting was no greater than in the AllMedium condition - resulting in the greatest accuracy of all of the conditions. It is the scaffolded sequence, from high constraint to medium to low that is beneficial, not just the variation in constraint: the Ascending condition resulted in less learning and greater forgetting.

While Study 2 uses an adult population, these results will answer questions about cognitive mechanisms in contextual word learning, which can extend to continuing work with younger populations. The use of the intelligent tutoring system will allow easy deployment to continue asking further questions about the ideal presentation of structured contextual word learning.

Combined, Study 1 and Study 2 provide insight into ways to support word learning at both the preliterate and literate stages. The results of both studies will inform futures studies on word learning, using both gesture and written context.
11 REFERENCES


Craik, F. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic


Frishkoff, G. A., Collins-Thompson, K., Crossley, S., & Hodges, L.E. (under review). An adaptive tutoring system for contextual word learning: The role of context cues and real time
feedback based on measures of partial knowledge.


Iverson, J. (2010). Developing language in a developing body: the relationship between motor
http://doi.org/10.1017/S0305000909990432

Iverson, J., Capirci, O., & Caselli, M.C. (1994). From communication to language in two modal-

Iverson, J., Capirci, O., Volterra, V., & Goldin-Meadow, S. (2008). Learning to talk in a gesture-
rich world: Early communication in Italian vs. American children. *First Language, 28*(2),
164-181. doi:10.1177/0142723707087736


ving adolescent literacy: Effective classroom and intervention practices: A Practice Guide*
(NCEE #2008-4027). Washington, DC: National Center for Education Evaluation and
Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Re-

Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with
cognitive tutors. *Educational Psychology Review, 19*(3), 239-264. doi: 10.1007/s10648-
007-9049-0

Koedinger, K. R., Booth, J. L., & Klahr, D. (2013). Instructional complexity and the science to


Namy, L. (2001). What's in a name when it isn't a word? 17-month-olds' mapping of nonverbal symbols to object categories. *Infancy, 2*(1), 73-86. doi: 10.1207/s15327078in0201_5


Özçalıskan, Adamson, Dimitrova. (submitted). Early deictic but not other gestures predict later vocabulary in both typical development and autism. *Autism Research.*


Van Daalen-Kapteijns, M., Elshout-Mohr, M., & de Glopper, K. (2001). Deriving the meaning of


Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit formation. *Journal of Comparative Neurology and Psychology, 18*, 459-482. doi:10.1002/cne.920180503
