A Comparison Between Nonlinguistic Cognitive Processing Treatment and Traditional Language Treatment for Bilingual Children with Primary Language Impairment

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Kerry Danahy Ebert

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Dr. Kathryn Kohnert

July 2011
Acknowledgements

I would like to thank my doctoral advisor, Dr. Kathryn Kohnert, for her guidance and support not only on this project but also throughout my doctoral program. Her mentorship is extraordinary and much appreciated. I would also like to thank the faculty members who have served on my preliminary and final exam committees, including Dr. Mary Kennedy, Dr. Benjamin Munson, Dr. Frank Symons, and Dr. Jennifer Windsor. They have each contributed to the development of this project and to my development as a researcher. My doctoral student cohort has also been an invaluable inspiration and support; I thank Dr. Kristina Blaiser, Dr. Kelly Nett Cordero, Dr. Pui Fong Kan, Amelia Medina, Dr. Giang Pham, Bita Payesteh, Jill Rentmeester, and Christine Wing. I am also thankful for the mentorship of Janet Jacobs.

I am truly appreciative of all of the people who contributed to making this project happen. The collaboration of the Minneapolis Public School district has been essential. I am particularly grateful for the efforts of Dr. Frank Cirrin; the district speech-language pathologists who assisted with identifying participants and generously loaned space and materials; and the principals, teachers, and staff at Andersen United, Armatage Montessori, Green Central, Jefferson, and Whittier Schools. I would also like to acknowledge the project personnel, many of whom volunteered their time and energy for this project. Jill Rentmeester and Bita Payesteh have put in countless hours serving as clinicians and project coordinators. Giang Pham has delivered both treatment and testing. Nicole Root has also administered treatment for the project. In addition, numerous important tasks were accomplished by our testers and laboratory assistants:
Maura Arnoldy, Shana Banas, Sheila Cina, Bao Dang, Mary DeChamps, Jasmine Egli, Christina Heinzen, Meg Introwitz-Williams, Daniel Jakab, Megan Jirschele, Phyllip Johnson, Jennifer Johnston, Laura Kawatski, Malka Key, Angie Luther, Olivia Matthys, Andrea Morales, Bethany St. Martin, and Christine Wing. Finally, I am grateful for our participants and families. Their enthusiasm truly made this project possible.

Funding for this project was provided by a Doctoral Dissertation Fellowship from the Graduate School of the University of Minnesota and by a grant from the National Institute on Deafness and Other Communication Disorders (1R21DC010868, titled *Three Treatments for Bilingual Children with Language Impairment* and awarded to Kathryn Kohnert). Additional funding sources that contributed to preliminary studies include the Bryng Bryngelson research fund through the Department of Speech-Language-Hearing Sciences at the University of Minnesota, the Graduate Research Partnership Program Summer Fellowship through the Graduate School of the University of Minnesota, and the New Century Scholars Doctoral Scholarship through the American Speech-Language-Hearing Foundation. Finally, the funding I received through the Leslie E. Glaze Scholar Award allowed my PhD program to take shape.

Last but certainly not least, I would like to thank my family. They have both given me the tools I need to succeed and sacrificed their own needs to support me as I worked towards this goal. I am very fortunate.
Dedication

This dissertation is dedicated to Matt. I am humbled by his steadfast support for me in all my endeavors.
Abstract

**Background**: Children with Primary Language Impairment (PLI) show subtle weaknesses in nonlinguistic cognitive processing (NCP) skills such as attention, memory, and speed of processing. It is possible that these weaknesses contribute causally to the language delays that characterize PLI. For bilingual children with PLI, NCP weaknesses would underlie language learning ability for both languages. The purpose of this study was to explore the relationship between NCP skills and language skills in bilingual children with PLI by treating processing speed and attention.

**Methods**: A total of 24 participants in three groups (NCP treatment, English language treatment, and delayed treatment control) completed the study protocol. All participants were Spanish-English bilingual children with PLI between the ages of 6 and 10 years. All participants completed an extensive battery of assessments indexing NCP, English language, and Spanish language skills both before and after a treatment cycle. Analyses examined change for individuals, for each group separately, and for the three groups in comparison to one another.

**Results**: Children who completed the NCP treatment showed significant change in processing speed and in overall English language skills. However, children who completed the English language treatment tended to make greater gains, both in English and in NCP skills. Few comparisons between the three groups reached significance, in part because the delayed treatment control group tended to make positive change and in part because of the small sample size. Individual variability was apparent across all three groups, but particularly pronounced for Spanish.
Conclusions: The results support a connection between NCP and language skills in children with PLI. Language-based treatment programs may effectively alter NCP skills, and NCP treatment programs may alter language skills.
# Table of Contents

Acknowledgements ........................................................................................................... i

Dedication........................................................................................................................ iii

Abstract............................................................................................................................ iv

Table of Contents ............................................................................................................ vi

List of Tables.................................................................................................................... x

List of Figures.................................................................................................................. xi

Chapter 1: Introduction..................................................................................................... 1

Chapter 2: Review of Literature ....................................................................................... 3

  Speed of Processing in Children with PLI ........................................................... 4

  Treatment of processing speed ..................................................................... 6

  Attention Skills in Children with PLI ................................................................. 8

  Approaches to attention............................................................................... 9

  Treatment of attention skills...................................................................... 15

  Memory Skills in Children with PLI ................................................................. 17

  Bilingual Children with PLI ....................................................................... 20

Chapter 3: Preliminary Studies.......................................................................................... 24

  Auditory Sustained Attention Task ............................................................. 24

  Pilot testing and results............................................................................. 26

  NCP Treatment for Monolingual Children with PLI ........................................ 26

  Results....................................................................................................... 28

  NCP Treatment for Bilingual Children with PLI ......................................... 28
Results. ......................................................................................................................... 30

Chapter 4: Research Questions .................................................................................. 32

Chapter 5: Methods ..................................................................................................... 35

Participants ................................................................................................................. 35

Assessment Measures ................................................................................................. 37

Nonlinguistic cognitive measures ............................................................................. 38

Spanish and English language measures ................................................................ 45

Treatment Conditions ................................................................................................. 49

NCP treatment. .......................................................................................................... 50

EL treatment. ............................................................................................................. 53

Sites .............................................................................................................................. 56

Clinicians .................................................................................................................... 57

Procedures .................................................................................................................. 58

Recruitment and screening ........................................................................................ 59

Group assignments ..................................................................................................... 59

Pre-testing sessions ................................................................................................... 60

Intervention sessions .................................................................................................. 61

Post-testing sessions ................................................................................................... 62

Participant Attrition ................................................................................................... 63

Treatment Fidelity ....................................................................................................... 64

Analyses ....................................................................................................................... 68

Individual change. ...................................................................................................... 68
Chapter 6: Results ........................................................................................................... 78
  Individual Change ........................................................................................................ 78
    NCP assessments .................................................................................................... 78
    English language assessments ............................................................................. 80
    Spanish language assessments ............................................................................ 81
  Absolute Change ....................................................................................................... 84
    NCP assessments .................................................................................................... 84
    English language assessments ............................................................................. 85
    Spanish language assessments ............................................................................ 86
  Relative Change ....................................................................................................... 87
    NCP assessments .................................................................................................... 87
    English language assessments ............................................................................. 87
    Spanish language assessments ............................................................................ 88
  Treatment satisfaction ............................................................................................... 89
  Post-hoc Analyses ..................................................................................................... 90
    Age ....................................................................................................................... 90
    Attendance .......................................................................................................... 90

Chapter 7: Discussion ..................................................................................................... 92
Revisiting Research Questions ........................................................................... 92
  Effects of treatment on NCP skills. ............................................................ 92
  Effects of treatment on English language skills. ........................................ 97
  Effects of treatment on Spanish language skills................................. 101

Additional Considerations ............................................................... 105
  Age. .......................................................................................................... 105
  Common factors. ...................................................................................... 106

Future Directions ....................................................................................... 110
  Time......................................................................................................... 110
  Task refinement. ....................................................................................... 111
  Individual differences. ............................................................................ 112

Chapter 8: Summary and Conclusions ......................................................... 114

Tables ............................................................................................................... 116

Figures ............................................................................................................... 134

Bibliography ...................................................................................................... 136

Appendices ......................................................................................................... 153
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>Participant characteristics and results from pilot testing of sustained attention task</td>
<td>116</td>
</tr>
<tr>
<td>Table 2</td>
<td>Repeated dependent measures and results from preliminary SSED studies</td>
<td>117</td>
</tr>
<tr>
<td>Table 3</td>
<td>Nonlinguistic cognitive processing measures administered at pre- and post-testing</td>
<td>118</td>
</tr>
<tr>
<td>Table 4</td>
<td>English and Spanish language measures administered at pre- and post-testing</td>
<td>119</td>
</tr>
<tr>
<td>Table 5</td>
<td>Participant characteristics and pretesting scores by treatment condition</td>
<td>120</td>
</tr>
<tr>
<td>Table 6</td>
<td>Description of treatment groups</td>
<td>122</td>
</tr>
<tr>
<td>Table 7</td>
<td>Treatment fidelity summary</td>
<td>123</td>
</tr>
<tr>
<td>Table 8a</td>
<td>Individual change on English language measures</td>
<td>124</td>
</tr>
<tr>
<td>Table 8b</td>
<td>Individual change on Spanish language measures</td>
<td>126</td>
</tr>
<tr>
<td>Table 8c</td>
<td>Individual change on nonlinguistic cognitive processing measures</td>
<td>127</td>
</tr>
<tr>
<td>Table 9a</td>
<td>Results of Fisher matched-pairs exact test comparing pre-test to post-test scores for nonlinguistic cognitive processing treatment group</td>
<td>128</td>
</tr>
<tr>
<td>Table 9b</td>
<td>Results of Fisher matched-pairs exact test comparing pre-test to post-test scores for English language treatment group</td>
<td>130</td>
</tr>
<tr>
<td>Table 9c</td>
<td>Results of Fisher matched-pairs exact test comparing pre-test to post-test scores for delayed treatment group</td>
<td>131</td>
</tr>
<tr>
<td>Table 10</td>
<td>Results of Aligned-Rank procedures comparing three groups on each outcome measure</td>
<td>132</td>
</tr>
</tbody>
</table>
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Summary of procedures across time</td>
<td>134</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Percentage improvement across languages for each participant</td>
<td>135</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction

Children with primary language impairment (PLI) show nonlinguistic cognitive processing (NCP) weaknesses in areas such as attention, memory, and processing speed. The relationship between these subtle weaknesses and the obvious language deficits in PLI is not yet clear. However, it is possible that the NCP weaknesses may contribute causally to language impairment. If the weaknesses lead to less efficient language learning ability or less efficient language processing, treating these weaknesses could help to remediate the obvious deficits in language.

Benefits of such a treatment could be particularly important for bilingual children with PLI. These children are overwhelmingly treated solely in English despite possessing a communicative need for proficiency in two languages. In contrast, treating underlying cognitive processing deficits could lead to gains in both languages. Bilingual children also present an ideal population for testing the hypothesis that cognitive processing weaknesses can lead to less efficient language learning ability, because comparisons between gains in each language can provide insight into the mechanisms of language change.

The purpose of this study is to evaluate the outcomes of a brief, intensive, NCP treatment for Spanish-English bilingual children with PLI. The outcomes for the experimental treatment are compared to outcomes of a more traditional linguistic treatment conducted solely in English and to outcomes of no treatment. Outcomes measured are the cognitive processing skills hypothesized to contribute to language learning, as well as language skills in both English and Spanish. It is hypothesized that
children treated with the NCP treatment will show greater gains in targeted cognitive processing skills than children in the traditional treatment group. It is also hypothesized that these gains will lead to improvements in both Spanish and English language skills. Whether these hypotheses are upheld or rejected, study outcomes have the potential to increase understanding of the relationship between cognitive and language skills in children with PLI.
Chapter 2: Review of Literature

Though they have traditionally been defined by the presence of impaired language skills alongside typical nonverbal cognitive skills (e.g., Tomblin et al., 1997), children with PLI demonstrate a host of weaknesses in nonlinguistic cognitive skills. These weaknesses do not result in clinically impaired nonverbal intelligence scores and their relationship to the language deficits apparent in PLI is a source of controversy (cf. Kohnert & Ebert, 2010; Leonard et al., 2007; Ullman & Pierpont, 2005; van der Lely, 2005). However, one possibility is that the language deficits in PLI are caused, at least in part, by subclinical nonlinguistic processing weaknesses.

The range of these subtle nonlinguistic weaknesses is substantial, though they do not appear in all areas (see reviews in Ullman & Pierpont, 2005; and Kohnert, Windsor, & Ebert, 2009). Children with PLI have demonstrated relative deficits in speed and accuracy while completing a variety of nonlinguistic tasks, including mental arithmetic (Fazio, 1999; Windsor, Kohnert, Loxtercamp, & Kan, 2008), auditory pattern matching (Yim, Kohnert, & Windsor, 2005), shape rotation (Windsor et al., 2008), and processing rapidly changing sensory inputs (e.g., Tallal et. al., 1996; Schul, Stiles, Wulfeck, & Townsend, 2004). In contrast, they may have relatively typical skills in orienting attention to visuospatial stimuli (Schul et al., 2004).

Two broad areas of deficit that may have particular implications for intervention are speed of processing and attention skills. Both areas have received substantial consideration in the literature, may underlie language deficits in PLI, and may be improved through training. These characteristics make speed of processing and
attention prominent candidates for improvement through nonlinguistic cognitive intervention.

The purpose of this chapter is to discuss the literature that supports treating speed of processing and attention skills in bilingual children with PLI. I will first consider the evidence for slowed processing in PLI, along with the evidence that this slowed processing contributes causally to PLI and the evidence that slowed processing can be improved with behavioral treatment. Next, I will review the evidence for subclinical impairments in the attention skills of children with PLI. This leads into consideration of how attention may be a mechanism of change in some language treatment studies, and how attention can be successfully treated through behavioral activities. I will also describe weaknesses in the PLI population in a third cognitive area, memory. Memory is not a focus of the experimental treatment, but may be affected by either language or NCP treatment. Finally, I will describe the population of bilingual children with PLI and discuss reasons why this population may be the ideal recipients of treatment targeting attention and speed of processing.

**Speed of Processing in Children with PLI**

Children with PLI are slower to complete a variety of both linguistic and nonlinguistic tasks (e.g., Kohnert & Windsor, 2004; Miller, Kail, Leonard, & Tomblin, 2001; Windsor & Kohnert, 2004). They are not as fast as their typically developing (TD) peers to name pictures, to mentally rotate shapes, to scan visual arrays, or to distinguish auditory words from nonwords. Their slowed performance on nonlinguistic tasks, which cannot be explained by their clinically apparent language deficits, has
sometimes been interpreted as evidence of a general information processing weakness (see Windsor, 2002, for a review).

In addition, simulating a slower speed of information processing in TD children has been shown to induce PLI-like performance. Hayiou-Thomas, Bishop, and Plunkett (2004) administered a grammaticality judgment task to 6-year-old children with typical language skills. When the participants’ processing capacities were stressed because the speech signal was presented at a faster rate, they made errors that mirror those of English-speaking children with PLI on this type of task: they had difficulty detecting errors in verb morphology but not errors in noun morphology. The results of Hayiou-Thomas et al. (2004) provide support for the idea that slowed cognitive processing speed could cause the language deficits seen in PLI.

A factor analysis by Leonard et al. (2007) also provides support for a causal role for slowed processing in PLI. A large group of 14-year-old children with and without PLI completed an array of linguistic and nonlinguistic tasks in addition to standardized language tests. Leonard et al. (2007) used the results to develop models that could predict the language skills of the participants. The best models included both linguistic and nonlinguistic processing speed factors, providing additional support for the idea that nonlinguistic processing speed is one factor that underlies the language deficits in PLI. The authors also indicate that their results point to the utility of intervention strategies that extend beyond language-specific goals. In particular, it may be important to address information processing of both linguistic and nonlinguistic information in treatment for PLI.
Treatment of processing speed.

Thus, there is notable support for the presence of slower information processing in children with PLI and for its contribution to language deficits. However, this evidence is not sufficient to justify treatment of processing speed. In order for interventions that target processing speed to be successful, it must also be a modifiable skill. To date, processing speed has not been trained in children with PLI (aside from the data published in Ebert and Kohnert, 2009, and presented below).

However, a growing body of literature suggests that processing speed can indeed be trained in a variety of other populations, including typical adults, aging adults, and children with disabilities. Evidence compiled from six studies which included over 2,000 older adults with impairments in processing speed indicates that training can produce immediate and lasting improvements in processing speed (Ball, Edwards, & Ross, 2007). Training programs in these studies were described as “computer-based, nonverbal exercises that are presented very briefly and involve target detection, identification, discrimination, and localization” (Ball et al., 2007, p. 20). The benefits of these programs lasted at least 2 years and brought about improvements in daily activities in addition to faster task performance.

Additional evidence that processing speed can be modified comes from studies of the effects of fast-paced action video games (Dye, Green, & Bavelier, 2009). Data from several studies have indicated that people who routinely play action video games have faster processing speeds – as measured by reaction time (RT) on visual processing tasks – than those who do not, without a corresponding drop in accuracy or evidence of
impulsivity. Furthermore, faster reaction times can be trained in people who do not play video games through a program of video game training. Dye et al. (2009) report data from 25 non-video game players who participated in 50 hours of training over 8 to 9 weeks. Participants trained on an action video game demonstrated a 13% drop in RT for visual search and discrimination tasks. In contrast, participants who were trained on a control (non-action) video game demonstrated a smaller RT improvement of 6%.

Finally, gains in speed of processing have been achieved through training in a few populations of children with disabilities. Yildirim, Erbaçeci, Ergun, Pitetti, and Beets (2010) reported improvements in reaction time for children with mild intellectual disabilities following a structured exercise program. However, processing speed in this study was measured using reaction time on a simple detection task, which may be more reflective of perceptual-motor speed than cognitive processing speed (Kohnert & Windsor, 2004).

Another instance of training processing speed in children with disabilities is reported by Cosper, Lee, Peters, and Bishop (2009). A small group of children with Attention-Deficit/Hyperactivity Disorder (ADHD) plus comorbid developmental coordination disorder or pervasive developmental disorder was trained using the Interactive Metronome program (Cosper et al., 2009). The 15-week Interactive Metronome program was hypothesized to train attention and motor coordination; however, the training produced improvements in reaction time rather than accuracy on a Continuous Performance Test (conventionally used to measure sustained attention).
Cosper et al. (2009) concluded that the training program was effective in treating reaction time but not sustained attention.

The concept of training people to perform cognitive tasks faster is thus gaining momentum in the literature. Though the training programs, dependent measures, and populations have differed in each study, the evidence is sufficient to support the hypothesis that speed of processing can indeed be improved with treatment.

**Attention Skills in Children with PLI**

Attention is often assumed to be impaired in children with PLI, though until recently evidence to support this hypothesis was relatively sparse. Some evidence that attention skills are weak in children with PLI comes from the high rates of comorbidity between PLI and the clinical disorders of attention, Attention Deficit Disorder (ADD) and ADHD. For example, Beitchman, Hood, and Inglis (1990) found that 34% of boys and 37.5% of girls with language disorders also met criteria for ADD. More recently, McGrath et al. (2008) found a 39% comorbidity rate between children with both language and speech disorders and ADHD.

However, not all children with PLI have attention deficits that reach clinical levels. Like processing speed, attention may be impaired at subclinical levels in children with PLI and these subclinical deficits may contribute to impairments in language learning. Attention can be conceptualized in many different ways, and the underlying conceptualization drives the way attention is assessed and attention impairment is defined. As such the evidence for these subclinical attention impairments
in children with PLI is best considered along with a description of the major conceptualizations of attention.

**Approaches to attention.**

Attention is a complex cognitive skill with diffuse effects. It has the potential to underlie learning on a variety of tasks. Because of these features, attention has been studied from different perspectives and several distinct approaches to characterizing attention have been developed. Here I will briefly consider three different approaches to attention, along with the theoretical and empirical reasons to suspect a link between attention and PLI that are associated with each approach.

**Controlled.**

In one approach to attention, the distinction between controlled and automatic information processing is central (Shiffrin & Schneider, 1977). Controlled processing involves conscious devotion of resources to a task, whereas automatic processing occurs without awareness. Automatic processing does, however, have the power to interrupt controlled attention to a task (Shiffrin & Schneider, 1977). This process may play a role in language learning, as a language learner must employ controlled attention to focus on relevant auditory input, and bar interruption from other processes.

The historic distinction between these two modes of attention is maintained in some current work on attention (e.g., Kane, Bleckley, Conway & Engle, 2001; Marton, 2008). This line of work uses higher-level cognitive tasks, such as the Tower of London and Wisconsin Card Sorting tasks, to assess controlled attention. In addition, controlled attention, working memory, and executive functions are all tightly linked
within this approach (Kane et al., 2001; Miyake, Friedman, Shah, Rettinger, & Hegarty, 2001).

Finally, this line of work suggests that children with PLI have difficulty in controlling their attention (Marton, 2008), as evidenced by difficulty in switching rules and inhibiting prepotent responses on tasks such as the Wisconsin Card Sorting task. However, the relationship between controlled attention and PLI may be mediated by a common association with working memory; working memory is a well-supported area of weakness in children with PLI (Kohnert, Windsor, & Ebert, 2009).

**Limited Capacity.**

A second approach to working with attention is to characterize it as a limited capacity system (Montgomery, Evans, & Gillam, 2009). According to this viewpoint, individuals have a fixed resource capacity for processing information. They can allocate these resources differently according to the attentional requirements of the task at hand, but cannot exceed their internal capacity. Breakdowns in attention will occur when task demands exceed that capacity.

Attention is again very closely linked to working memory according to this view, and can even be seen as a subcomponent of working memory (Montgomery et al., 2009). Thus it is perhaps unsurprising (again, given the link between working memory impairments and PLI) that children with PLI have been found to have poorer attention skills than TD peers from the limited capacity viewpoint of attention as well (Montgomery et al., 2009).

**Component Models.**
A third way to conceptualize attention is to break it down into parts. Such an approach allows a more complete characterization of the points of breakdown in impaired attention, and also accounts for neuroanatomical evidence (Mirsky et al., 1991). The component approach to attention encompasses more than one model. Specific components differ among models. For example, Gomes et al. (2000) divide attention into four major subcomponents: arousal, orienting, selective attention, and sustained attention. Arousal refers to alertness and the readiness to perceive stimuli; orienting is an automatic process that occurs when a novel stimulus is detected; selective attention is an active process of focusing on a desired stimulus; and sustained attention is the ability to maintain focus over time (Gomes et al., 2000).

Alternatively, Mirsky et al. (1991) propose a distinct – though overlapping – set of four components (focus, sustain, shift, and encode). Kerns, Eso, and Thomson (1999) suggest that most models include several basic components:

- the ability to sustain attention over time (vigilance), the ability to attend to stimuli selectively, the ability to alternate or switch attention stimuli or tasks [sic], and the ability to divide attention so as to maintain more than one ongoing process. (p. 274)

Recent evidence suggests that, in comparison to TD peers, children with PLI have deficits in some areas of attention but not others. In particular, sustained attention in PLI has been the target of growing interest in the literature (e.g., Finneran, Francis, & Leonard, 2009; Spaulding, Plante, & Vance, 2008). In order to synthesize the evidence for subclinical sustained attention deficits in the PLI population, Ebert and Kohnert (in
press) conducted a meta-analysis of studies comparing the performance of children with and without PLI on tasks of sustained attention. The prototypical assessment of sustained attention is the continuous performance task (CPT, Mirsky et al., 1991). In a CPT – known also as a go/no-go, monitoring, or vigilance task – participants are asked to detect target stimuli from a stream of auditory or visual stimuli. The results of the meta-analysis supported the existence of PLI deficits on these tasks, in comparison to TD peers. When 28 effect sizes from 15 studies were combined, the resulting weighted average Hedge’s $g$ effect size was 0.69 (Ebert & Kohnert, in press). This result differed significantly from zero ($Z = 13.17, p < 0.001$). In other words, children with PLI demonstrate significant differences from TD peers on sustained attention tasks, falling on average nearly three-quarters of a standard deviation below these peers.

The meta-analysis (Ebert & Kohnert, in press) also addressed the question of whether sustained attention deficits in PLI are modality-specific. Some authors have argued that these deficits are limited to the auditory domain (e.g., Noterdaeme, Amorosa, Mildenberger, Sitter, & Minow, 2001; Spaulding et al., 2008). However, the meta-analysis did not support domain-specificity, when the linguistic content of stimuli are considered. That is, when tasks that use linguistic auditory stimuli are separated from those that use nonlinguistic auditory stimuli, significant differences in weighted average effect size exist only between visual tasks and auditory linguistic tasks (not between visual tasks and auditory-nonlinguistic tasks). In addition, the weighted average effect size, $g$, is significantly different from zero across all three categories (auditory-linguistic: $g = 0.82, Z = 11.34, p < 0.001$; auditory-nonlinguistic: $g = 0.61; Z$
Thus, it appears that children with PLI perform more poorly on sustained attention tasks when they are asked to process language-based information than nonlinguistic information, as we might expect from children defined by language delays. Sustained attention deficits in children with PLI do not appear to be limited to the auditory domain.

There is evidence that divided attention may also be impaired in children with PLI. Tropper, Marton, Russo-Victorino, Shafer, and Schwartz (2008) tested children aged 10 to 12 years on a divided attention task. Children were asked to decide if an image matched a word presented to one ear while ignoring a word presented to the other ear. Ignored words varied in their relationship to the image; some of the words to be ignored matched the image whereas others were unrelated. TD children followed patterns of performance that could be predicted based on the expected level of interference from the ignored words, whereas the performance of children with PLI deteriorated in the presence of any distracting word, regardless of its relationship to the image. This result was interpreted as a deficit in divided attention.

Niemi, Gunderson, Leppäsaari, and Hugdahl (2003) also used a divided attention task to differentiate participants with and without PLI. Niemi et al. (2003) found that five family members with PLI were not as successful as control participants at focusing attention to consonant-vowel syllables heard through the left ear while ignoring syllables heard through the right ear.

Selective attention is implicated as well by the evidence for PLI deficits in sustained and divided attention. The ability to focus on a specific target -- the
conventional definition of selective attention -- is a component of both sustained attention tasks and divided attention tasks. In fact, divided attention is viewed as a subcomponent of selective attention in some component models (Gomes et al., 2000). Some authors attempt to differentiate sustained attention from selective attention by comparing performance at the end of a CPT from performance at the beginning, but this distinction has not been supported well by the literature on PLI (Finneran et al., 2009; Noterdaeme et al., 2001). Because of the close relationship between these subcomponents of attention, it is common to refer to the CPT and similar assessments as sustained selective attention tasks.

Finally, preliminary evidence suggests that children with PLI do not differ from typically developing peers in their attentional arousal or orienting skills. Noterdaeme et al. (2001) found no differences between children with and without PLI on the alertness task of the attentional battery that was administered. Schul et al. (2004) found that school-aged children with PLI did not differ from peers in their ability to orient to novel visuospatial stimuli.

In summary, there is substantial literature to support subclinical impairments in the attention skills of children with PLI. Evidence of impairment has been found from all three viewpoints discussed here, and in multiple areas within the component approach. Sustained selective attention appears to be particularly impaired in PLI, based on current evidence. Because of this evidence as well as the close relationship between attention and other cognitive constructs such as working memory within the
controlled and limited capacity viewpoints, the remainder of this study will consider attention primarily from a component view with a focus on sustained selective attention.

However, it is important to note that the literature surrounding attention and its links to PLI is not yet as developed as the literature on speed of processing in PLI. There are no studies to date that manipulate attention to induce PLI (as Hayiou-Thomas et al., 2004, did for speed of processing) or look at attention as a factor in language performance (as Leonard et al., 2007, did for speed of processing). Instead, the primary argument for treating attention directly in children with PLI comes from prior treatment studies that have argued for a role of attention in language gains.

**Treatment of attention skills.**

In addition to studies that have assessed attention skills in children with PLI, it has been speculated that language interventions, and particularly computerized language interventions, produce language gains in part via treatment of selective and sustained attention (Ebert & Kohnert, 2009; Gillam, Crawford, Gale, & Hoffman, 2001; Gillam et al., 2008; Stevens, Fanning, Coch, Sanders, & Neville, 2008). However, these investigations have lacked direct behavioral measures of attention. Stevens et al. (2008) reported that children with PLI demonstrate event-related potential (ERP) patterns consistent with deficits in selective attention prior to treatment and that these deviant patterns were reduced following intensive computerized language intervention. Language skills also improved after the intervention. Stevens et al. (2008) interpreted the results to indicate that the computerized intervention had improved selective
attention and thereby language skills, although they did not include behavioral measures of attention in the study.

Both Gillam et al. (2001) and Gillam et al. (2008) compared multiple intervention programs for school-aged children with PLI. Both studies also used intensive intervention schedules. Each study found similar results across the programs that were compared and in both cases the authors speculated that children in different conditions may have experienced common gains in sustained selective attention skills as a result of intensive practice processing visual and auditory stimuli. These information processing gains may have contributed to the observed gains in language skill. A similar conclusion is drawn by Ebert and Kohnert (2009; see Preliminary Studies for more detail). Again, attention skills were not directly measured in any of these three studies. Thus, while there are promising suggestions that language interventions – particularly those that are intensive and at least partially computer-based – may promote gains in attention, direct evidence of these gains is not yet available.

Studies in other populations indicate that sustained and selective attention skills can indeed be trained. Bavelier and colleagues have developed a substantial literature (e.g., Dye & Bavelier, 2010; Green & Bavelier, 2003) indicating that playing action video games develops visual selective attention in typical children and adults. Their studies indicate that overall attentional capacity in the visual domain is better in video game players than in non-video game players across a variety of tasks. In addition, when non-video-game-playing adults are trained on action video games, their performance improved significantly on visual selective attention tasks. Bavelier and
colleagues argue that the need to manage multiple tasks at once in action video games enhances attentional capacity.

In children with disabilities, attention skills have also been successfully improved through treatment, and specifically through cognitive treatments. For example, 38 school-aged children with acquired brain injuries due to trauma, infection, or malignancy participated in a program to improve attention through daily exercises and strategic behavioral modification (van’t Hooft et al., 2007). The participants showed significant improvements in sustained attention, as measured by a CPT, in comparison to a control group. In addition, the gains were maintained over 6 months.

Similarly, cognitive training has been used to improve attention in children with ADHD (Kerns et al., 1999). Kerns et al. (1999) treated seven children with ADHD using “process-specific activities”; that is, games requiring focus on specific auditory or visual features. Children that participated in this training improved more than children who received a nonspecific treatment on a number of measures of selective and sustained attention. However, there was no change in behavioral indices of attention.

In summary, sustained and selective attention, like speed of processing, appears to be a trainable skill in a variety of populations, including children with impairments. Furthermore, it is possible that intensive language interventions are one means of training attention. However, behavioral evidence of this is lacking to date.

**Memory Skills in Children with PLI**

In addition to processing speed and attention, a third major area of cognitive weakness exists in the PLI population: memory. Like attention, memory can be
conceptualized in different ways. Memory weaknesses in PLI are usually characterized as either short-term memory deficits (e.g., Hayiou-Thomas et al., 2004) or working memory deficits (e.g., Montgomery, 2002), depending on the viewpoint.

Evidence for subtle PLI deficits in short-term or working memory is substantial in both the linguistic and nonlinguistic domains. One of the most widely used measures of linguistic working memory is nonword repetition (NWR). In an NWR task, a child must retain and repeat a nonsense word. Poor performance on NWR tasks is one of the best-documented cognitive-linguistic weaknesses in the PLI profile. Graf Estes, Evans, and Else-Quest (2007) found a weighted average effect size of 1.27 across 23 studies that compared PLI and TD performance on English NWR tasks, indicating that on average children with PLI performed more than 1 standard deviation below their TD peers.

Evidence of PLI weaknesses in short-term or working memory extends to nonlinguistic tasks as well. For example, children with PLI have greater difficulty than TD peers in recalling spatial locations (Hoffman & Gillam, 2004). Children with PLI also perform below TD peers in counting and remembering digits (Danahy, Windsor, & Kohnert, 2007).

Short-term or working memory deficits have also been linked causally to language deficits in PLI. Hayiou-Thomas et al. (2004) were able to induce PLI-like errors in TD children by increasing memory load in the same way that they simulated PLI by increasing processing speed load. Working memory emerged alongside processing speed as one of the two main factors leading to PLI language performance in
Leonard et al.’s (2007) factor analysis. Thus, both of these studies singled out speed of processing and memory as cognitive factors underlying language deficits in PLI.

However, memory is not the primary focus of the experimental treatment program in this study. There is conflicting evidence as to whether short-term or working memory can be improved with training. Traditional models of working memory (e.g., Just & Carpenter, 1992) suggest that individual working memory capacity is fixed. The efficiency of operations in processing, storing, and retrieving information can be improved, leading to better performance on working memory tasks. However, this type of improvement might be more closely related to processing speed gains than memory gains per se.

In contrast, some of the recent literature in this area has proposed that training on working memory tasks does increase memory performance. For example, Buschkuehl et al. (2009) trained healthy 80-year-old adults on a working memory task for three months. Improvements were seen in overall memory performance, but particularly in visual working memory.

In my own preliminary work in treating NCP skills in children with PLI (Ebert & Kohnert, 2009; described in Chapter 3), I did not find compelling evidence of memory improvements. This result, along with the theoretical controversy over the plasticity of memory capacity, led to a shift in focus in the NCP treatment program, away from memory and towards attention and processing speed. However, short-term or working memory remains an established NCP deficit in PLI, and will be considered in this study as a potential area of change following treatment.
Bilingual Children with PLI

As the number of foreign-born residents of the United States increases, an increasing number of children in the U.S. are raised in homes in which a language other than English is spoken (Schmidley, 2001). The home language becomes the first language (L1) for the child. These children typically learn English as a second language (L2) when they enter the school system, thus becoming sequential bilinguals. When bilingual children are affected by PLI, both L1 and L2 are negatively impacted (Kohnert, 2007).

It is not yet established how often sequential bilingual children are affected by PLI (Kohnert & Medina, 2009). In the monolingual population, PLI affects approximately 7% of school-aged children (Tomblin et al., 1997), but the same type of epidemiological research has not yet been conducted in bilingual populations in the US. It is likely that PLI is a high-incidence disorder for bilingual children, as it is for children who speak only English. Given the approximate rates of both bilingualism and PLI in the US, it is logical that thousands of bilingual children with PLI reside in the US.

Unfortunately, appropriate assessment and treatment of these children presents a substantial challenge to speech-language pathologists (Caesar & Kohler, 2007; Kohnert, Kennedy, Glaze, Kan, & Carney, 2003). The vast majority of speech-language pathologists serving bilingual populations are themselves monolingual; for example, 98% of respondents in the survey conducted by Caesar and Kohler (2007) were monolingual clinicians with bilingual children on their caseloads. In addition, few
respondents used interpreters (Caesar & Kohler, 2007). Thus speech-language pathologists are easily able to provide intervention services in English but are challenged to provide the same support in the home language.

However, it is highly likely bilingual children with PLI will require the active support of both home and community languages (e.g., Kohnert, Yim, Nett, Kan, & Duran, 2005). The very limited evidence available on treatment outcomes for bilingual children with PLI shows advantages for bilingual treatments over single-language support (see Kohnert & Medina, 2009, for review). To date, there is little evidence that children with PLI are able to independently transfer skills learned in English to their home language. In addition, even typically developing bilingual children may have difficulty learning language skills taught only in English (Kohnert & Danahy, 2007; Roseberry & Connell, 1991). In short, intervention services only in English are likely to provide inadequate support to a bilingual child with PLI (Restrepo, 2005; Kohnert, 2007).

One solution to this problem is to work with family members and interpreters to provide intervention services in both languages. Another potential method of producing gains in both languages may be to improve underlying cognitive processing skills. Bilingual children with PLI demonstrate the same cognitive processing weaknesses as their monolingual counterparts; they perform below bilingual peers with typically developing language skills on nonlinguistic processing speed and memory tasks (Kohnert, Windsor, & Pham, 2009). If treatment of these weaknesses can lead to language improvements in one language for monolinguals, it may also lead to
improvements in both languages for bilinguals. As yet, nonlinguistic cognitive treatments have not been examined in bilingual children with PLI (aside from Ebert, Kohnert, & Rentmeester, 2011; discussed in the Preliminary Studies section). There is some evidence that treating underlying cognitive skills – such as simple arithmetic, categorization, and visual scanning – can lead to gains in both languages in a bilingual adult with aphasia, an acquired primary language impairment (Kohnert, 2004).

Bilingual children with PLI also constitute a unique population for testing hypotheses about the relationship between nonlinguistic cognitive and language skills. In monolingual children with PLI, improvements in language skill following treatment with NCP tasks may be attributed at least in part to extraneous factors such as social interaction in the language of intervention. In contrast, bilingual children possess an additional area of potential improvement (i.e., L1) that provides a comparison for gains made in the language of intervention (presumably, L2). If cognitive processing skills underlie successful language acquisition and use, improvements in these skills should transfer to both languages in a bilingual child, provided that opportunities to use both languages remain available outside the treatment setting. Some differences in the magnitude of gains in L1 and in L2 following NCP treatment might be expected, given the differences in opportunities for language use between the home language and the school language. However, a complete lack of gains in L1 would suggest the mechanism of change differs from the hypothesis that change in nonlinguistic processing skills leads to language gains. Instead, the social interactions in L2 at
treatment sessions and the setting of treatment at school, where L2 is primarily used, might play a greater role in language gains than originally hypothesized here.
Chapter 3: Preliminary Studies

Three preliminary studies contributed to the development of one assessment task and to the experimental treatment activities. These studies provide background for methodological choices and also provide preliminary support for the use of NCP treatment in children with PLI. The procedures and results of each preliminary study are described below.

Auditory Sustained Attention Task

Because the purpose of the present study is to examine changes in both NCP skills and language skills, assessment tasks that accurately index skills in each area of interest are crucial. Established tasks for assessing nonlinguistic speed of processing and memory were available (see Table 3 for details). However, it was necessary to develop a measure of sustained selective attention.

The sustained selective attention task that we developed was based upon literature reviewed in the meta-analysis of sustained attention (Ebert & Kohnert, in press). More specifically, two recent high-quality studies (Finneran et al., 2009; Spaulding et al., 2008) created sustained attention tasks in the CPT format that separated children with PLI from TD peers. Spaulding et al. (2008) used noise-degraded environmental sounds, presented in 96 sequences of 3. Finneran et al. (2009) presented 200 shapes very briefly (400ms). Both tasks were sensitive to the PLI-TD difference, but with children younger than the participants in the current study.

We chose to develop a sustained attention task using environmental sounds because of the possibility that auditory tasks are slightly more sensitive for detecting
sustained attention deficits in the PLI population (see Ebert & Kohnert, in press). In addition, the assessment of processing speed – the other nonlinguistic cognitive skill targeted by the experimental treatment – uses visual stimuli. Creating an auditory attention task prevented the NCP assessments of interest from being limited to a single sensory domain.

Like Spaulding et al. (2008), we chose environmental sounds related to vehicles as stimuli. Four sounds became the distracter sounds: a car horn, a car door handle opening, a race car engine revving, and a car engine starting. The fifth sound, jingling keys, was selected as the target sound. All sounds were obtained from the free online directory of sound files provided by the Comparisonics Corporation (www.comparisonics.com).

Unlike Spaulding et al. (2008), we did not degrade the sounds by adding noise. We chose instead to increase task difficulty by using brief stimuli, as in Finneran et al. (2009). Sounds were edited to last 500 milliseconds (ms), with an additional 500 ms available for the child to respond. The interval before the next sound was set at 500 ms. Thus, the complete cycle for one stimulus lasted 1500 ms. To create a task that lasted 5 minutes, 200 stimulus cycles were included. Finally, the percentage of target sounds was set at 20%, creating 40 targets and 160 distracters.

Dependent variables included both overall accuracy and RT. Ebert and Kohnert, in press, found evidence for sustained selective attention impairments in children with PLI using accuracy measures from CPTs but not using RT measures. However, a small
number of studies examined RT, and we chose to measure it in order to build more
evidence in this area.

**Pilot testing and results.**

This auditory sustained selective attention task was piloted with 14 children in
May 2010. The purpose of pilot testing was to establish task feasibility and difficulty
levels within the target age range. The pilot testing group was heterogeneous with
respect to the presence of both language and attentional impairments, and with respect
to bilingual status. A summary of both participant characteristics and results from pilot
testing for the attention task can be found in Table 1.

Although the pilot testing demonstrated smooth task administration to the target
population, results showed substantial ceiling effects. Eleven of fourteen children
obtained overall accuracy scores of at least 195 out of 200. In other words, the original
task did not appear sufficiently difficult to detect subtle attention difficulties, or to allow
room for improvement as a result of treatment.

In response to these results, the difficulty of the original task was increased by
doubling the task length. This adjustment resulted in a 10-minute task with 400 stimuli.
All other task parameters were left unchanged.

**NCP Treatment for Monolingual Children with PLI**

The NCP treatment protocol was developed and piloted through two treatment
studies. Both studies used single-subject experimental design (SSED) to establish the
linguistic and nonlinguistic effects of treating cognitive processing skills using
nonlinguistic stimuli. In the first study (Ebert & Kohnert, 2009), two monolingual
English-speaking girls with severe PLI completed an intensive intervention designed to improve processing speed and memory skills. The study followed a multiple baseline design. Two distinct cognitive areas, visual processing speed and auditory memory, were targeted by the intervention, and participants completed repeated dependent measures designed to index linguistic and nonlinguistic skills in both areas.

Thus, there were four repeated dependent measures used to assess the constructs of interest. A choice visual detection task (Kohnert & Windsor, 2004) assessed nonlinguistic visual processing speed. To assess linguistic processing speed, a Rapid Automatic Naming (RAN) task was created. The two auditory memory assessments were a nonword repetition task (based on Frisch, Large, & Pisoni, 2000) in the linguistic domain and an auditory pattern matching task (Yim et al., 2005) in the nonlinguistic domain. Brief descriptions of these four assessments and study outcomes for each measure appear in Table 2. In addition to the four repeated dependent measures, pre- and post-testing with standardized language assessments was included as a more traditional measure of language proficiency.

Participants completed a four-week intervention, with four 90-minute intervention sessions scheduled each week. The introduction of treatment to each area was staggered across participants to create four baselines of varying length. All six of the NCP intervention activities used in the current study were also employed in this preliminary project. The two participants also completed two additional software games from the Earobics Step II software package (Earobics, 2006), designed to target auditory memory skills.
Results.

Results of the project suggested that the NCP activities had a positive effect on language performance in the participants. Both participants demonstrated large improvements in standardized scores for global language and expressive vocabulary, as measured by the standardized tests. Global language gains were driven primarily by improvements in sentence formulation and production of grammatical morphemes. As shown in Table 2, naming speed, as measured by the RAN task, improved notably during the intervention period for both participants. Finally, one participant demonstrated speed improvements on the choice visual detection task during intervention.

In summary, there was substantial evidence of change in visual processing speed and lexical access. In contrast, there was little support for change in auditory memory or receptive language. In response, the two Earobics games designed to target auditory memory were dropped from the NCP intervention protocol. Instead, emphasis in treatment and assessment was shifted away from auditory memory and towards attention. Attention emerged in this preliminary study as a potential mechanism of language change, but it was not directly assessed. The sustained selective attention measure was developed in response to this need.

NCP Treatment for Bilingual Children with PLI

A second SSED study (Ebert, Rentmeester, & Kohnert, 2011) was completed to replicate the previous findings, using the adjusted NCP treatment protocol and bilingual participants. Two Spanish-English bilingual boys, aged 8;4 and 7;5, participated in a
four-week intervention protocol. Like Ebert and Kohnert (2009), the study was designed to follow a multiple baseline design, with repeated measures tracking the skills of interest. However, this study eliminated the distinctions between targeted cognitive skills that existed in the previous study, and therefore compared multiple baselines across participants only (versus across participants and behaviors). Furthermore, attendance problems on the part of one participant resulted in baselines of the same length for both participants, limiting adherence to SSED principles.

Nonetheless, the study provided an opportunity to examine the effects of the treatment protocol with the population of interest. The treatment activities were identical to the NCP treatment activities in the current study. Thus the goal of treatment was to improve speed of processing and attention, with potential transfer to both Spanish and English. Treatment sessions were 90 minutes long and scheduled four days per week during summer school programming. One participant (P3 in Table 2) attended 13 of 14 scheduled treatment sessions, whereas the second participant (P4 in Table 2) attended only 7 of 14 scheduled treatment sessions. All treatment sessions were conducted entirely in English, with a monolingual English-speaking clinician.

Repeated dependent measures were similar to the previous SSED, but adjusted slightly based on the results of Ebert and Kohnert (2009) and the bilingual status of the participants. Participants again completed the choice visual detection task (Kohnert & Windsor, 2004) as a measure of visual processing speed. They also completed a nonword repetition task in both English (based on Dollaghan & Campbell, 1998) and in Spanish (based on Ebert, Kalanek, Cordero, & Kohnert, 2008). These sets of nonwords
were not as difficult as those used in Ebert and Kohnert (2009), which proved too
difficult for the participants and resulted in floor effects. The RAN measure was
retained. Finally, a sentence repetition measure was added as an additional measure of
auditory memory and language change. Brief descriptions of the repeated dependent
measures and study outcomes are found in Table 2. Standardized language assessments
in both Spanish and English were also administered in a pre- and post-testing format in
this study.

**Results.**

Evidence of change was found on almost all of the repeated dependent
measures, with the greatest change seen for both participants in Spanish and English
nonword repetition and in English sentence repetition. The choice visual detection and
RAN tasks provided weaker evidence of change, but nonetheless supported a case for
improvements in visual processing speed. Change on these two measures was also
consistent with the results of the previous SSED with the two monolingual English-
speaking children with PLI.

Standardized testing showed significant improvement in English morphosyntax
for one participant and in Spanish vocabulary for the other participant. Scores in other
areas were stable across pre- and post-testing; however, the treatment period was very
brief to produce changes on standardized language measures.

Overall, the results support exploration of the NCP treatment program on a
larger scale. It is particularly notable that Spanish language gains were obtained despite
a complete absence of Spanish interaction or treatment activities during intervention sessions.
Chapter 4: Research Questions

This study explores the relationship between cognitive processing skills, such as attention and processing speed, and language skills by attempting to change cognitive skills in children with PLI. More specifically, the purpose of this study is to evaluate the outcomes for bilingual children with PLI following an NCP treatment designed to improve sustained selective attention and processing speed. Two groups of children with PLI provide a comparison for outcomes from the NCP treatment: a no-treatment control group and a group receiving a traditional English-only language-based treatment for PLI.

Three areas form the outcomes of interest: the targeted NCP skills, language skills in L1, and language skills in L2. Within each of these three areas, three types of change can be considered: absolute change, or statistically significant change from pre-to post-testing within a group; relative change, or statistically significant change between the NCP treatment group and one or more of the comparison groups; and individual change, or clinically significant change for an individual, regardless of the group pattern. This study seeks to answer three multifaceted questions:

1. Does the NCP treatment produce improvements in sustained selective attention and in processing speed:
   a. for individuals?
   b. for the group?
   c. in comparison to no treatment and in comparison to an English language (EL) treatment?
Null hypothesis 1: NCP treatment does not significantly improve sustained selective attention or processing speed. A failure to reject the null hypothesis for this question would suggest that the NCP treatment is not effective in modifying the targeted cognitive skills.

2. What changes in L1 (Spanish) skills are apparent following the NCP treatment:
   a. for individuals?
   b. for the group?
   c. in comparison to no treatment and to EL treatment?

Null hypothesis 2: L1 skills do not change following the NCP treatment. A failure to reject the null hypothesis, assuming that NCP skills showed change and the first null hypothesis was rejected, would indicate that change in NCP skills does not always lead to change in language skills. Such a result could lend support to the hypothesis that nonlinguistic cognitive and linguistic skills are independent in children with PLI (e.g. van der Lely, 2005). In contrast, a rejection of this null hypothesis would support the hypothesis that some NCP skills underlie language performance in children with PLI (e.g., Leonard et. al., 2007).

3. What changes in L2 (English) skills are apparent following the NCP treatment:
   a. for individuals?
   b. for the group?
   c. in comparison to no treatment and to EL treatment?
Null hypothesis 3: L2 skills do not change significantly in response to NCP treatment. Like the previous null hypothesis, a failure to reject this null hypothesis would indicate changes in NCP skills do not necessarily lead to language changes. However, a rejection of null hypothesis 3 may need to be interpreted more cautiously, if null hypothesis 2 is not rejected; change in L2 (English) without change in L1 (Spanish) may indicate that peripheral aspects of the treatment, such as interaction with and attention from the examiners, play a role in observed language changes. Ongoing schooling in English and school-based speech-language services in English would also be expected to contribute to English language growth.
Chapter 5: Methods

Participants

A total of 24 Spanish-English bilingual children with PLI between the ages of 6;0 and 9;11 enrolled in the study. Of these children, 18 were male and 6 were female. All participants attended school in the Minneapolis Public School district and were receiving special education services for language disorder. Under the qualification criteria for special education services in the state of Minnesota, this designation indicates that each child scored at least 2 standard deviations below the mean on two separate language tests or demonstrated a substantial difference from expected performance on two nonstandardized measurement procedures (Minnesota Office of the Revisor of Statues, 2008). Given the inadequacy of standardized tests normed on monolingual children for diagnosing bilingual children, it is likely that the participants in this study qualified for language disorder services on the basis of nonstandardized measurement procedures (such as criterion-referenced instruments, language samples, and observations). In addition, the children did not have other primary disabilities to explain the language disorder, such as autism, cognitive delay, or other health impairments.

Thus all participants met conventional criteria for PLI according to the Minneapolis Public School district qualification process. Several steps were taken to verify that this qualification process had correctly diagnosed each participant. First, parent interviews were conducted to confirm that parents of participants were concerned with participants’ language development. Parental concern may be one of the clearest
indications of language impairment in bilingual children (Caesar & Kohler, 2007; Restrepo, 1998). Each parent was also asked to confirm the absence of hearing problems, autism, head injury, cerebral palsy, seizures, general cognitive delay, and physical problems in his or her child’s health history.

In addition, hearing screening and nonverbal intelligence testing were conducted at the first assessment session to ensure that participants did not have a hearing loss or cognitive delay that could explain their delayed language skills. All participants passed a hearing screening at 20dbHL at 1000, 2000 and 4000 Hz. All participants also completed the Test of Nonverbal Intelligence – 3rd Edition (TONI-3, Brown et al., 1997). A minimum standard score of 75 (1.66 standard deviations below the mean) on the TONI-3 was established prior to beginning the study (Tomblin et al., 1997). However, it was not necessary to exclude any participants from the study because of this requirement; the lowest standard score that any participant obtained on the TONI-3 was 81.

The bilingual status of participants was verified through school records, parent interviews, and intake testing. Most participants attended schools in which English is the sole language of instruction and received school-based language therapy services only in English. One of the five participating school sites, with two participating children, provided some Spanish language and literacy instruction as well as bilingual language therapy services. To determine home language use, parents of participants were interviewed about the languages spoken in the home. All parents indicated that they spoke Spanish either all the time or most of the time in the home. Parents were
also asked the age at which their children were first exposed to English; all participants were first exposed to English between birth and age four. In addition, all participants were able to complete both Spanish and English language testing in pre-test sessions, to the extent possible given the severity of the PLI.

Finally, because participants qualified on the basis of a diagnosis within the school system, all participants were receiving school-based services during their participation in the study. We judged it unethical to remove participants from their regular speech-language treatment in order to receive experimental treatment programs. Instead, the treatment provided by the study took place during after-school hours and supplemented school-based treatment. The existence of the delayed treatment group (which received school-based but not study-based treatment) and the difference in intensity between the school-based and study-based treatment (e.g., 60 minutes per week versus 360 minutes per week) served to separate the effects of the study treatment from the school-based treatment.

**Assessment Measures**

A battery of measures was selected to index targeted cognitive and linguistic skills both before and after treatment. These measures fall into three categories: non-linguistic cognitive, Spanish language, and English language. The assessment measures administered at pre- and post-testing are described below and summarized in Tables 3 and 4.
Nonlinguistic cognitive measures.

The nonlinguistic cognitive measures were designed to index the three skills most often suggested as areas of nonlinguistic cognitive weakness in PLI: attention, processing speed, and memory. In order to customize the measures to a specific cognitive skill and to use exclusively nonlinguistic stimuli, each measure was implemented as a nonstandardized, computerized task using the E-Prime software package (Psychology Software Tools, Inc., 2000).

Participants were trained on each measure before completing the full task. For each of the three tasks described below, training was completed prior to beginning the main task, and the main task was completed only if the participant understood the task. Details about the training are described on a task-specific basis.

In general, each measure generated an accuracy measure and a speed, or RT, measure. For all three tasks appropriate data trimming procedures were followed before calculating RT, to prevent outliers from confounding RT results. Task-specific procedures are described below.

Sustained selective attention.

A Continuous Performance Task was used to measure sustained selective attention. The development and piloting of this task is described in Chapter 3. The final task consisted of 80 instances of each of five sounds, in random order. The five sounds were all related to cars, with the sound of jingling keys serving as the target sound. Each sound cycle lasted 1500 ms: 500 ms of the sound; an additional 500 ms in which to respond; and 500 ms of interstimulus interval.
To complete the task, participants were seated in front of the computer. They were told they would hear some noises related to cars, and that they should push the response button when they heard the noise of keys. Each participant completed a practice version of the task with eight trials, two of which played the target sound. All 4 distracter sounds also played during the practice task. Two participants during the Cycle 1 pre-testing were unable to comprehend the task despite repeated instructions and practice. These two participants did not complete the main task during this session. All other participants proceeded from the practice task to the main task, which lasted 10 minutes.

Two primary measures were extracted from the resulting data to serve as dependent variables: $d'$ (d-prime) and RT. $d'$ is a signal detection measure that combines both hit rate and false alarm rate (Macmillan & Creelman, 1991). It measures the participant’s ability to detect the signal (in this case, the target sound) using standard deviation units. In other words, $d'$ serves as a single accuracy measure for detection tasks that takes into account any participant bias towards errors of commission (i.e., responding to a distracter sound) or omission (i.e., neglecting to respond to a target sound).

In order to calculate $d'$ for an individual, the individual’s hit rate and false alarm rate must be converted to Z scores. In cases where the individual obtains a perfect hit rate (i.e., 1.0) or false alarm rate (i.e., 0), a standard correction is applied to allow the Z scores to be calculated (Wixted & Lee, n.d.). The method adopted here estimates the true hit rate at $1 - 1/(2 \times \text{number of targets})$, the equivalent of missing $1/2$ of a hit, when a
perfect hit rate of 1 is obtained. Similarly, the false alarm rate is estimated to be \(1/(2 \times \text{number of distracters})\), the equivalent of assuming the participant had \(\frac{1}{2}\) of one false alarm, when the false alarm rate for the task is 0. Thus, in the 1 pre-testing and 3 post-testing instances in which the child correctly responded to all 80 targets, a hit rate of \(1-1/(2 \times 80) = 0.99375\) was used to calculate \(d'\); in the 6 pre-testing and 4 post-testing instances in which the child correctly did not respond to any of the 320 distracters (after correcting for late hits, as discussed below), a false alarm rate of \(1/(2 \times 320) = 0.0015625\) was used to calculate \(d'\).

The RT measure was the mean RT for all correct responses, after trimming any responses less than 50ms or more than 2 standard deviations from the individual mean RT. Note that in a CPT RT can be derived for target sound trials only, as there is no response to record on a correct distracter trial. Also, it was not necessary to set an absolute upper cutoff for RTs, because the maximum RT for this task was 1000ms. The task was programmed to play the next stimulus after that interval. Thus, if no response was received within the 1000ms window on the target trials, the trial became a miss.

This time limit has the potential to unfairly penalize slow responders by giving them an incorrect target trial, or miss, followed by an incorrect distracter trial, or false alarm, every time they respond too slowly to a target stimulus. In order to ensure that this scenario did not affect accuracy scores for the sustained selective attention task, all instances in which a miss was immediately followed by a false alarm, within 500ms, were marked as “late hits”. Accuracy for both the miss and subsequent false alarm was then manually adjusted for each late hit. The problem did not appear to be systematic;
these “late hits” constituted a total of 0.8% of all possible hits across all testing sessions, with a maximum of 6.3% of all possible hits in any one testing session.

**Processing speed.**

A choice visual detection task provided a measure of processing speed in the visual domain. Complete specifications for this task appear in Kohnert and Windsor, 2004. This task was also utilized in the two preliminary SSED studies (Ebert & Kohnert, 2009; Ebert et al., 2011). It was chosen to be the processing speed measure in the current study because it appears to be particularly sensitive to cognitive processing deficits in children with PLI (Kohnert & Windsor, 2004) and because of preliminary evidence that it may be sensitive to change after an intervention program (Ebert & Kohnert, 2009; Ebert et al., 2011).

To complete the choice visual detection task, participants sat in front of a computer with an attached response box. Participants were instructed to watch the screen for the appearance of a red or blue circle. Each participant was trained to use his or her index finger to push one response button for a red circle and the other response button for a blue circle. The need to respond as quickly as possible was emphasized during training as well as before beginning the main task. The practice task included 8 trials and the main task included 25 trials. One participant during Cycle 1 pre-testing was unable to comprehend and complete the choice visual detection task despite repeated instruction and practice.

The two dependent variables derived from this task were overall accuracy and RT. RT serves as the primary outcome measure because the construct of interest is
speed of processing; accuracy scores were obtained to ensure the validity of the RT measures. The RT score was defined as the mean for all correct trials after removing any RT less than 50ms, greater than 2000ms, or more than two standard deviations away from the individual mean. These trimming procedures resulted in the elimination of 5.4% of correct trials from RT calculations.

**Memory.**

The third cognitive task administered in this study, the auditory serial memory task, is designed to index auditory short term memory. Complete task details are available in Yim (2006). This task has been shown to be particularly sensitive to differentiating children with PLI from children without PLI in bilingual populations, especially when combined with the choice visual detection task (Kohnert, Windsor, & Pham, 2009). A shortened version of the auditory serial memory task constituted one of the repeated measures in the first preliminary SSED (Ebert & Kohnert, 2009), although it proved to be too difficult to be informative with those individual participants.

During the auditory serial memory task, the computer plays pairs of tone sequences. Sequences begin at two tones and become progressively longer during the task until they reach five tones each. Participants are instructed to decide whether the two sequences in the pair are the same or different. The frequencies of the tones composing the sequences are 250, 500, 1000, 2000, and 3000 Hertz. Tones last 100ms each with 500ms in between the tones. The computer screen shows a static picture of a person listening; to assist the participant in segmenting the auditory input into two sequences, the background behind the picture turns yellow during the first sequence of
each pair and blue during the second sequence. There is also a 500ms break between the first and second tone sequences.

Training for the auditory serial memory task had two phases. In the practice phase, participants completed 8 trials, with feedback for each individual trial. In the criterion phase, participants completed 8 trials without feedback. If a score of at least 6 out of 8 was not obtained, the practice and criterion phases were repeated. If the participant again did not obtain a passing score (i.e., 6/8) on the criterion phase, he or she did not progress to complete the main task. A number of participants fell into this category, including: 3 during Cycle 1 Pre-testing; 1 during Cycle 1 Post-testing; 2 during Cycle 2 Pre-testing; 1 during Cycle 2 Post-testing; 1 during Cycle 3 Pre-testing; and 0 during Cycle 3 Post-testing.

Participants who progressed to the main task completed a total of 66 trials, 6 of which were used for reliability purposes only. Thus there were 60 trials that counted towards each participant’s accuracy score for the task, or 15 at each of the four sequence lengths (i.e., two, three, four, or five tones per sequence). The accuracy score was the primary measure of interest for this task. RT was also measured. RT for the auditory serial memory task was calculated for correct “same” trials only. “Different” trials are excluded because it is possible to determine that the sequences are different before the end of the second sequence, causing the appropriate start time for measuring RT to vary by trial. In contrast, two sequences cannot be determined to be the same before the end of the second sequence, creating a constant starting point for measuring RT on “same” trials. RT outliers were also removed from the data before calculating an
individual’s mean. Removing outliers entailed excluding RTs less than 50ms or more than 10,000ms, and then excluding those RTs more than 2 standard deviations from the individual mean. The relatively large absolute upper cutoff for RTs (10,000ms, following Kohnert, Windsor, & Pham, 2009) was due to the task’s difficulty for this population and the resulting long and variable RTs. Data trimming procedures resulted in the elimination of 3.8% of trials from RT calculations.

Because of the wide range of performance on this task, neither the accuracy nor the RT measures alone were sensitive to change for all participants. As noted above, several participants were unable to comprehend and complete the task beyond chance accuracy; in contrast, several others scored near the maximum accuracy score at the pre-test, rendering RT a more sensitive dependent variable. However, an additional feature of the auditory serial memory task is an inherent hierarchy of difficulty. The task contains four levels of difficulty, corresponding to the four sequence lengths for tone stimuli. In other words, accurately completing the task for sequences with five tones is more difficult than completing the task for sequences of four tones, and four tones is more difficult than three tones and so on.

Therefore, a new dependent variable was derived to capture change across the range of participant performance, taking advantage of the difficulty hierarchy within the task. The new variable, called the achievement level, was defined as the highest level of task difficulty a participant could accurately complete. Achievement of a difficulty level was set at a score of at least 11 of 15 trials correct, with the additional condition that the same criterion be met for all lower levels of difficult. The standard of 11 of 15
trials correct was derived from several clinical speech-language pathology sources that suggest a skill is “acquired” at 75% accuracy (Agocs, Burns, De Ley, Miller, & Calhoun, 2006; Bleile, 1995; Boudreau & Hedberg, 1999). The validity of the standard is also supported by the infrequency with which participants met the standard for a given level without meeting it for all easier levels; this situation occurred only 3 times out of 45 potential instances.

For participants who were unable to pass the criterion phase or unable to reach 11 correct responses at the easiest level of difficulty (two-tone sequences), achievement level was 0. Participants who reached at least 11 correct responses for two-tone sequences received an achievement level score of 1. Participants that scored at least 11 of 15 for both two- and three-tone sequences received a score of 2. This scoring system was continued up to those participants who demonstrated achievement of all task levels. For these participants, a fifth level of task difficulty was created by considering RT scores. Because increasing speed while maintaining accuracy represents improvement in task performance (e.g., Fabrizio & Moors, 2003), participants who had scored at least 11 of 15 on all 4 task difficulty levels and also decreased their mean RT from pre- to post-testing by at least 15% of their pre-test mean RT were given an achievement level score of 5.

**Spanish and English language measures.**

All language measures employed in the study have two parallel versions: one in Spanish and one in English. The measures were designed to provide a comprehensive
profile of language skills across domains and languages. All language measures are
summarized in Table 4.

Three standardized tests were administered in each language: the Expressive
One-Word Picture Vocabulary Test (EOW-E, Brownell, 2000a) in English and the
Expressive One-Word Picture Vocabulary Test - Bilingual Edition (EOW-S, Brownell,
2001a) in Spanish provided measures of expressive vocabulary; the Receptive One-
Word Picture Vocabulary Test (ROW-E, Brownell, 2000b) in English and the
Receptive One-Word Picture Vocabulary Test – Bilingual Edition (ROW-S, Brownell,
2001b) in Spanish provided measures of receptive vocabulary; and the Core Language
score from the Clinical Evaluation of Language Fundamentals – 4th Edition in English
(CELF-4E, Semel, Wiig, & Secord, 2003) and in Spanish (CELF-4S, Wiig, Secord, &
Semel, 2006) provided measures of global language skills. The Core Language score
is composed of four subtests, but the subtests vary according to age. For children aged 5-
8 years, the Concepts and Following Directions, Word Structure, Recalling Sentences,
and Formulated Sentences subtests make up the Core Language score. For children
aged 9 years or older, the Concepts and Following Directions, Recalling Sentences,
Formulated Sentences, and Word Classes-Total subtests make up the Core Language
score.

The EOW-S and ROW-S were intended to be measures of Spanish vocabulary
only in this study. That is, each test was to be administered entirely in Spanish, with
only Spanish responses counted as correct, to provide a parallel comparison to English
vocabulary scores. However, the tests as originally published were designed to measure
bilingual composite vocabulary (i.e., to count either English or Spanish responses as correct). Due to a miscommunication, some research assistants administered the tests bilingually, according to publisher instructions, and others administered the tests in Spanish only, the modification intended for this study. For all group comparisons, only scores obtained solely in Spanish are reported. For individual analyses, change from pre- to post-testing was calculated only if the test was administered in the same manner on both occasions.

In addition to the standardized language measures, there were two nonstandardized language measures in both Spanish and English. First, the Rapid Automatic Naming task from the CELF-4E and CELF-4S was used as a measure of naming and linguistic visual processing speed. In the preliminary SSEDs (Ebert & Kohnert, 2009; Ebert et al., 2011), the RAN measure in English appeared highly sensitive to change from the NCP treatment. The RAN in this study was administered in a nonstandardized manner, rather than following the CELF-4 procedures; only the demonstrations and color-shape naming trial were administered to each participant.

Following the SSED results, the primary dependent variable of interest from the RAN was intended to be the time from the color-shape naming trial, measured in seconds. Accuracy, as measured by the number of errors, was also recorded. Recording accuracy was intended primarily to ensure that children did not sacrifice accuracy in order to achieve speed; in other words, we expected relatively high accuracy for all participants, allowing us to examine speed. However, not all participants demonstrated near-perfect accuracy on the RAN.
Two notable sources of error were cross-language intrusions and lack of color and/or shape vocabulary in Spanish. Cross-language intrusions occurred when a participant substituted a color or shape word from the language not being tested (e.g., calling a triangle “triángulo” on the English RAN) or substituted the typical word order for color-shape descriptions from the language not being tested (e.g., saying “circle green” instead of “green circle” on the English RAN). In addition, several participants had difficulty completing the Spanish RAN because they could not fluently name colors and shapes in Spanish. The reverse was not observed; that is, all participants could fluently name colors and shapes in English. Colors and shapes are academic vocabulary that is more likely to be familiar to a bilingual child in the school language (in this case, English) than in the home language. Children who could not fluently name colors and shapes in Spanish were not asked to complete the Spanish RAN beyond the trial items. Five participants did not complete the Spanish RAN at pre-testing and five participants did not complete the task at post-testing.

The number of errors made during the RAN for those children who did complete it ranged from 0 to 36. In order to capture both speed and accuracy in RAN scores, the ratio of the completion time in seconds to the number of accurate responses was used as the primary dependent measure for this task.

The second nonstandardized language measure is an NWR task. The NWR task was administered by playing recorded instructions and stimuli over headphones. Participants were asked to listen carefully to the words and repeat them as accurately as possible. The English stimuli consist of 16 nonwords ranging in length from one to four
syllables. The Spanish stimuli consist of 20 nonwords ranging in length from one to five syllables. A list of stimuli and a complete description of their development is found in Dollaghan and Campbell (1998) for the English nonwords and in Ebert et al. (2008) for the Spanish nonwords.

Responses were digitally recorded and later scored by trained research assistants. Scoring procedures followed Dollaghan and Campbell (1998); deleted and substituted phonemes were counted as incorrect, whereas added phonemes and distorted phonemes were not. The number of correct phonemes was then totaled and divided by the total number of phonemes to create a percent phonemes correct (PPC) score.

Each audio file was scored by at least one research assistant. During the first treatment cycle, all NWR audio files were scored by two research assistants, with disagreements resolved by discussion to reach consensus scoring. This procedure was followed in order to ensure that all NWR coders were trained to be consistent. In addition, 15% of audio files across Cycles 1, 2, and 3, in both Spanish and English, were randomly selected for reliability scoring. These files were re-scored by this author, and compared on a phoneme-by-phoneme basis to the original scores. Overall interjudge reliability following this procedure was 88.2%.

Treatment Conditions

Three treatment conditions are contrasted in this study: experimental NCP treatment, traditional EL treatment, and delayed treatment. The two active treatment groups (experimental NCP and traditional EL) were designed to parallel each other as much as possible in the structure of treatment and the types of treatment activities as
well as in the skill of the interventionist and the intensity of the treatment delivery. As such, treatments differed only in the content of the activities. Each treatment used three interactive games and three computer games. Each treatment and its respective activities is described below.

**NCP treatment.**

The experimental treatment group received the NCP treatment program piloted in Ebert and Kohnert (2009) and Ebert et al. (2011). The focus of this treatment is improving speed of processing and attention skills, and the content of all six selected activities is designed toward that end. All activities use nonlinguistic or minimally linguistic stimuli; in games with multiple settings, clinicians were specifically instructed to use only the settings that were nonlinguistic. For example, some of the visual stimuli were shapes, colors, and abstract symbols; some of the auditory stimuli were tones and musical noises.

Clinicians were also instructed to avoid using any typical linguistic treatment techniques while interacting with children. Such techniques include recasts, expansions, and corrections of child utterances. However, clinicians were allowed and encouraged to interact normally with the children for all procedural aspects of treatment, such as greeting participants, providing instructions, feedback and encouragement, and transitioning between activities.

**NCP Computer Games.**

All three computer games came from the Locutour Multimedia Attention and Memory: Volume II software package (Scarry-Larkin & Price, 2007). The three
selected programs -- Change, Scanning, and Dominoes -- were all used successfully
during the pilot treatment studies. The Change game requires rapid decision making
using changing visual information in a go/no-go task format. The child watches a series
of visual stimuli appear on the screen, and is instructed to respond to target stimuli only.
The target stimulus changes multiple times during the course of one game. For
example, the stimulus stream might consist of triangles and circles, with circles as the
target stimulus. The child would respond to circles and ignore triangles until instructed
to change, and then he or she would ignore circles and respond to triangles. Participants
were also encouraged to respond as quickly as possible.

In the Scanning game, participants must find a set of target symbols in
increasingly complex visual arrays. For example, the participant might see an array of
35 teardrop shapes in different colors, and be asked to find and click on all of the blue
ones. The program provides a steady stream of visual arrays, with a changing target for
each, throughout the activity block. The activity is timed and participants are instructed
that they must find all of the targets as quickly as possible.

In the Dominoes game, participants must recognize matching tiles, rotate them,
and align them. In more difficult levels, basic arithmetic skills are required to identify
matches. This game requires higher-level cognitive skills than Change or Scanning, as
participants compete against the computer and can achieve higher scores through
strategy and planning. Each turn is timed; if participants do not place a tile quickly
enough they forfeit the turn.
Three interactive games were also used as experimental treatment activities. The first game, Blink (Staupe, 2001), is a card game. Participants compete with each other or with the clinician to play all of their cards as quickly as possible. Cards can only be played if they match the previously played card in the shape, number, or color of images on the card. Thus, Blink requires participants to sort images rapidly and flexibly.

The second game, Bop-It, is a handheld device that uses musical noises to give motor commands. The participant tries to complete as many actions as possible. The time between noises becomes smaller as the game continues, increasing the difficulty. The game requires auditory attention to process the musical noises quickly enough to complete the actions.

Two versions of Bop-It were used in the study: Bop-It Blast and Bop-It Extreme. Both versions fit the description above; they differ only in that Bop-It Blast has three musical noises and associated actions whereas Bop-It Extreme has five. The initial treatment plan called for Bop-It Blast only, but Bop-It Extreme had to be used as well due to the failure of a Bop-It Blast device and its lack of commercial availability.

The third game, Simon Trickster, is also a handheld device. It requires participants to replicate tone and light sequences of increasing length, taxing both visual and auditory attention to task.
EL treatment.

Participants assigned to the traditional EL treatment group received an English language-based treatment involving a combination of commercially available language therapy activities. The focus of treatment is expanding vocabulary, practicing morphosyntactic constructions, and following directions to improve language comprehension. Each activity is explicitly designed to target at least one of those three language goals.

Like the experimental NCP treatment, the traditional EL treatment includes three computer software activities and three interactive activities. The activities were chosen based on a combination of the existing literature with regard to treating children with PLI and the need to mimic the structure of the experimental treatment. The work of Gillam and colleagues (Gillam et al., 2001; Gillam et al., 2008) has indicated that intensive intervention using Laureate Learning Systems and Earobics software (the two software packages employed in this study) is effective at improving language. In contrast, there is little support in the literature for non-computerized language intervention programs in this age group. A recent systematic review of language interventions for school-aged children with PLI (Cirrin & Gillam, 2008) as well as a comprehensive manual of language treatment approaches (McCauley & Fey, 2006) reveal few empirically tested intervention approaches for the 6;0 to 9;11-year-old age group that do not employ computerized activities or target phonological awareness or written language. These skills are not targets of the traditional treatment program in this study. Cirrin and Gillam (2008) cite some support for semantically-based
vocabulary treatment in this age group (Wing, 1990), which is targeted by the selected games.

Finally, one potential language intervention program, the narrative-based intervention employed in the recent large-scale treatment study conducted by Gillam and colleagues (2008), was not selected because it did not result in significant gains over the control condition in the published study and because its format might lead to differences in participant interest between treatments in this study. Clear differences in the participants’ interest in the NCP and EL treatment activities would potentially confound results. Therefore, because of the lack of support for other programs and the need to maintain structural similarity between the treatments, three games were chosen to complement the three computer games in the EL treatment program.

Because only one of the six activities had been utilized in one of the preliminary studies, some pilot testing was indicated. Limited pilot testing for the EL treatment activities was carried out in June 2010. The purpose of the pilot testing was to ensure that task difficulty levels were generally appropriate for the target population and that the games were sufficiently complex to hold children’s attention over a seven week treatment period. With one exception (described in detail in the next section), the pilot testing indicated that the EL treatment activities met these criteria.

*EL computer games.*

Two of the three computer games are published by Laureate Learning Systems. Adjectives and Opposites (Wilson & Fox, 1997) teaches descriptive vocabulary words.
Participants are presented with circus scene and asked to identify the new vocabulary, which includes adjectives and pairs of opposites.

The second game published by Laureate Learning Systems was originally scheduled to be Prepositions (Wilson & Fox, 2008). Prepositions trains comprehension of ten prepositional phrases that describe spatial relationships. However, pilot testing indicated that this game would not be sufficiently difficult for older participants. Therefore, an additional computer game titled, Swim, Swam, Swum, was purchased from Laureate Learning Systems to serve as a complement to Prepositions. As with Prepositions, Swim, Swam, Swum targets morphosyntactic skills, one of the main focus areas of the EL treatment program. However, Swim, Swam, Swum teaches high-frequency irregular verbs, rather than prepositional phrases.

The third computer game, Calling All Engines (Earobics, 2006), requires participants to follow directions of increasing length and complexity. Participants are instructed to find one, two, three, or four items, in the correct order, with or without a delay between the instruction and its execution. The game automatically adjusts difficulty level based on the child’s performance.

**EL interactive games.**

Three interactive games were also used in the traditional treatment program. All three selected games are produced by LinguiSystems, Inc. Category Card Games (Anderson, 2004) addresses vocabulary growth through semantic organization. It consists of cards with images of items belonging to several different semantic categories (e.g., sports, body parts, rooms in a house, letters, etc.). Participants played games with
the cards that required them to identify categories, describe other items in the category, or sort cards by category.

Gram’s Cracker: A Grammar Game (Cole, 2000) is the second interactive game in the traditional treatment program. Gram’s Cracker targets multiple morphosyntactic constructions, such as pronouns, plurals, possessives, negatives, and verb tenses. Participants are taught the constructions in both receptive and expressive language modalities.

The final game is Plunk’s Pond: A Riddles Game for Language (LinguiSystems, 1998). Plunk’s Pond is designed to stimulate vocabulary growth and improve listening skills. The clinician reads the participant a series of clues that describe an object or action. The participant is asked to deduce the described word.

Sites

This study took place via a partnership with the Minneapolis Public School District. All study procedures took place in one of five participating elementary schools: Andersen United, Armatage Montessori, Green Central, Jefferson, and Whittier. To participate in the study, children enrolled in the afterschool program at their own elementary school, and study procedures took the place of afterschool programming. Because the NCP treatment was considered experimental, we did not consider it appropriate to remove participants from any portion of their regular academic instruction.
The Minneapolis Public School District is an urban district serving a diverse population of students. Across the district, 26%\(^1\) of students are identified as English Language Learners or ELLs. Within the five schools participating in the study, this figure varies widely, from 3% of students at Armatage Montessori to 59% of students at Anderson United. The five schools also vary in the percentage of students qualifying for free or reduced lunch (a commonly used indicator of low income). At Armatage, 30% of students qualify; at Whittier, 61% qualify; at Jefferson, 82% qualify; at Green Central, 89% qualify; and at Anderson United, 93% of students qualify for free or reduced lunch. The participating schools are more homogeneous with respect to the percentage of students qualifying for special education services. Four of the five schools fall within a few percentage points of the district-wide average of 15% for this figure: Anderson (15%), Armatage (14%), Green (14%), and Whittier (16%). At Jefferson, 24% of students receive special education services. Finally, Jefferson participates in the Native Language Literacy program, in which some classroom instruction (including literacy instruction) is provided in Spanish. The remaining four school sites provide instruction only in English.

**Clinicians**

Four clinicians administered the treatments. All clinicians had an earned Master’s degree in speech-language-hearing sciences and disorders (or its equivalent), were certified by the American Speech-Language-Hearing Association and had multiple years of experience working as a speech-language pathologist with pediatric

\(^{1}\) All figures in this paragraph are based upon statistics for the 2009-2010 school year.
populations. Three of the clinicians were doctoral students in speech-language pathology, and the remaining clinician was an employee of the Minneapolis Public School District. Details on the therapy groups and the clinician who administered each one appear in Table 6.

A separate group of individuals administered all pre- and post-testing measures. Due to the amount of time needed for pre- and post-testing, a total of 17 individuals assisted with the testing. This group included 9 students pursuing masters’ degrees in speech-language pathology; 1 student pursuing a clinical doctorate in audiology; 3 students enrolled in undergraduate programs in the Department of Speech-Language-Hearing Sciences; 3 doctoral students with clinical certification in speech-language pathology; and 1 faculty member with clinical certification in speech-language pathology. All examiners were trained on the assessment tasks before administering them to participants. All examiners who administered tasks in Spanish were required to be fluent Spanish speakers.

**Procedures**

To complete the study, participants proceeded through multiple stages, including screening, group assignment, pre-testing, treatment, and post-testing. All procedures took place at the participants’ school sites. Complete study procedures involved attendance at approximately 30 sessions over 8-9 weeks. Each session lasted 80 to 100 minutes, the length of the afterschool program. A summary of the study phases and the order of participants’ progress through them can be found in Figure 1.
Treatment procedures were repeated in three cycles: one in Fall 2010, one in Winter 2011, and one in Spring 2011. These time periods are referred to as Cycles 1, 2, and 3, respectively, throughout this manuscript. All participants completed the study procedures during one of the three treatment cycles.

**Recruitment and screening.**

Participants were recruited through speech-language pathologists in the Minneapolis Public School District. Speech-language pathologists were provided with approved recruitment flyers and consent forms and asked to distribute them to students on their caseloads who fit the profile of interest. Families were subsequently contacted in Spanish via phone to ensure that they understood the forms and to answer any questions they might have. Interested parents then returned signed consent forms to the school, where study personnel collected them.

Potential participants were then screened through phone calls to parents. The phone screening ensured that each potential participant spoke both Spanish and English; that s/he had no identified psychosocial, neurological, or hearing disorders; and that the participant’s parents were concerned with his or her communication skills. If the participant met all study criteria, s/he was enrolled and assigned to a group. Parents were asked to make certain that their child was enrolled for afterschool programming.

**Group assignments.**

Participants were placed into groups of three according to their school site. In schools with more than three enrolled participants, assignment to groups was done
randomly. Each group of three was then assigned to receive the NCP or the EL treatment.

In addition, a delayed treatment control group was created. Participants assigned to the delayed treatment group did not attend any of the study treatment sessions between the pre- and post-testing sessions. Like the participants receiving active intervention, they did continue to participate in school-based services during the study. After participation in the delayed treatment group, children were assigned to receive one of the active study treatments during a subsequent treatment cycle. For example, children in the delayed treatment group during Cycle 1 received one of the active treatments during Cycle 2 or Cycle 3. Assignment to the delayed group was not random due to logistical and ethical constraints.

The groups were compared on all pre-testing variables using the Kruskal-Wallis test of ranks (Kruskal & Wallis, 1952) to ensure there were no significant differences between groups. Table 5 contains group mean scores for each of the background and pre-testing variables, as well as the results of the Kruskal-Wallis test for each variable.

**Pre-testing sessions.**

The purpose of the pre-testing sessions was to index selected language and cognitive processing skills prior to intervention. All assessment measures described above were administered during the pre-testing sessions. The initial schedule specified two pre-testing sessions, with all Spanish language tasks administered in one session and all English language tasks administered in the other. However, due to logistical issues (e.g., absences, additional time needed to complete tasks for some participants)
actual procedures deviated somewhat from this schedule. It was not possible to complete all assessment measures in two sessions for the majority of participants. Instead, pre-testing took place over two to four sessions. Whenever possible, English and Spanish testing were not conducted in the same session. If it was necessary to mix English and Spanish within the same assessment session, the languages were still separated by examiner (i.e., one examiner carried out Spanish tasks and another carried out English tasks within the session).

**Intervention sessions.**

Participants in the two active intervention groups were originally scheduled to complete 28 sessions of treatment over 7 weeks. However, unplanned cancellations to after-school programming resulted in fewer than 28 sessions available in each treatment cycle. The exact number of treatment sessions conducted ranged from 22 to 24, depending on the afterschool programming for the site and cycle.

Within each session, 75 minutes were devoted to intervention activities. The remaining minutes (out of approximately 100 minutes per session) were devoted to transitions between activities and break time. The exact length of transitions and break time varied somewhat depending on school site and treatment program.

Five 15-minute activity periods made up the direct intervention time for both active treatment groups. Participants rotated through a different activity in each of the five activity periods. Treatment schedules ensured that each child played each game in the treatment program with roughly equal frequency. Treatment schedules were also structured such that interactive games were sometimes played one-on-one with the
clinician for individual feedback, and sometimes played with another participant in addition to the clinician to provide peer modeling and interaction.

**Post-testing sessions.**

In the post-testing sessions, participants repeated all tests given in the pre-testing session in order to measure change in language and cognitive processing skills during the experimental period. Sessions were again divided by the language of test administration as much as possible. Post-testing required 2 to 4 sessions per child.

In addition, both participating children and their parents were asked to complete a treatment satisfaction survey as a qualitative, additional outcome measure. The surveys were administered only to participants in the two active treatment groups. The purpose of these surveys was to assess client and family beliefs about the treatment programs and to document client and family experiences with the treatment programs. Client belief in treatment efficacy may play a substantial role in treatment outcomes (e.g., Ebert & Kohnert, 2010; Wampold, 2001). English and Spanish versions of the parent and child satisfaction surveys are shown in Appendix A and B, respectively. The parent satisfaction survey was adapted from a similar measure used in a treatment study conducted by Rvachew and Nowak (2001), and then translated into Spanish. In this study, a Spanish-speaking liaison contacted participants’ parents to complete the satisfaction survey. The parent satisfaction survey is reproduced in both Spanish and English in Appendix A.

To complete the children’s version of the survey, two of the clinicians who administered treatment met with participating children after the completion of treatment
(e.g., at a post-testing session). The questions were read to each child in English, with the child either answering verbally or marking his or her answer on the paper. Survey administration was scheduled such that the clinician who administered the survey to any given child was not the same clinician who had administered that child’s treatment.

The children’s satisfaction survey is included in Appendix B.

**Participant Attrition**

A total of 24 consent forms were signed. When children who participated in two cycles of the study (e.g., first as a control participant and subsequently in an active treatment group) are counted once for each cycle, participation across the three cycles totals 28. There were three instances in which a child did not complete the study in the treatment group to which he or she was initially assigned. One child completed partial pre-testing for Cycle 1 but discontinued participation for unknown reasons. This child returned and completed the study during the next treatment cycle. Another child had a change in school schedule that prevented him from continuing to participate, but he remained a control participant for that cycle.

One child completed assessment procedures, but was determined to have a severe speech sound disorder that impacted his speech intelligibility to a degree that language testing was not deemed reliable. Separate from the current study, this child was provided with treatment for the speech sound disorder, in Spanish. His speech intelligibility improved and he completed all study procedures in the next treatment cycle.
In addition, four participants completed the study but their data were excluded from analyses due to inconsistent attendance at treatment sessions. Three of these children were assigned to an NCP treatment group and one was assigned to an EL treatment group. These children attended 9 to 13 treatment sessions. In contrast, all participants that were included in analyses completed at least 16 treatment sessions, or at least 75% of the scheduled sessions in the treatment cycle to which they were assigned. Reasons for poor attendance for the four excluded children included child illness, family travel, and fluctuating interest on the part of the child or family in the optional after-school programming offered at each school.

**Treatment Fidelity**

To ensure that both active treatments were implemented as intended, videotapes of approximately 19% of intervention sessions were reviewed. All fidelity reviews were conducted by this author. The purpose of the fidelity review was to verify compliance with the treatment schedule and the intended activities for both active treatments. The specific variables extracted from each recorded session were the length of each activity and the transition between activities; the name of each activity; and a count of the number of comments made by the clinician in each of four categories: Spanish, unrelated, redirect, and traditional linguistic cuing (in English). Each variable was extracted on a participant-specific basis. For example, the number of times the clinician directed English linguistic cues to each participant was counted, as was the amount of time each participant spent on breaks and transitions. A summary of the data from the fidelity review is shown in Table 7.
Operational definitions of each of the four types of comments were established prior to beginning fidelity review. Spanish comments were defined as any utterances in the Spanish language, occurring at any point in the treatment session. The purpose of counting comments in this category was to verify that children were not receiving Spanish stimulation during treatment sessions. The fidelity review indicated that during one EL treatment session, a clinician gave the Spanish translation of a stimulus word. This single-word utterance was the only Spanish stimulation noted for either of the treatment groups in any session reviewed.

Unrelated comments were defined as any utterance that was not for the purpose of directing the child’s attention to the treatment activity or directly related to the activity at hand. These comments were tallied only during activity periods. For example, the following exchange would count as an unrelated comment for the clinician, if it occurred during a treatment activity:

Child: “I have a brother.”

Clinician: “really? what’s his name?”

A different response from the clinician, such as, “we can talk about that later. Let’s do our work now”, would not have counted as an unrelated comment, because it redirects the child to the treatment activity. Tallying unrelated comments ensured that clinicians were primarily using the prescribed activities as the means of delivering treatment, rather than engaging excessively in extraneous conversations with participants. The mean number of unrelated comments per session was 0.7 for the NCP sessions and 1.3 for the EL sessions. All unrelated comments that were observed were
in response to a participant statement or question during a treatment activity. Thus, the clinicians occasionally answered an off-task question, but overall remained focused on the treatment activities.

A “redirect” comment was counted each time the clinician cued the participant to focus on the activity at hand in response to an off-task participant behavior. For example, if a participant stepped away from his or her computer program during an activity period and the clinician responded with, “You need to do your work,” a redirect comment was counted. General encouragement and praise (e.g., saying, “you’re doing so well today! Keep it up!”) was not counted as redirecting. This category was tallied as a measure of participant engagement in the treatment activities, across both individuals and treatment programs, because attention to task was one of the constructs of interest in this study. As expected, the amount of redirection provided by the clinicians varied widely across participants and sessions, from a low of 0 redirects to a high of 53 (for a single participant in one session). However, the number of redirects did not vary systematically between the two groups, \( t(75) = 1.2, p = 0.22, d = 0.30 \).

Finally, the number of times the clinician provided linguistic cuing was counted. Linguistic cuing was defined to be any instance of a semantic or syntactic recast, expansion, or correction of a child comment. These comments were counted during both treatment activities and transitions. The primary purpose of this category was to ensure that clinicians administering the NCP treatment were not using language-based treatment techniques in addition to the NCP activities. The linguistic cues that were
observed during the NCP sessions tended to be clarifications of a child’s statement, such as the following exchange:

Child: “I make him win.”

Clinician: “You helped him win?”

The clinician here supplies a more appropriate verb and tense, thus providing the child with a linguistic cue.

During EL sessions, linguistic cues were not counted for stimulation that was built-in to the treatment activities (such as a computer command or the clinician reading a clue from a card). However, any additional linguistic cuing the EL clinician provided surrounding the treatment activities was counted. Though the definition of linguistic cuing was the same for both NCP and EL treatment sessions, linguistic cuing during the EL sessions went beyond the conversational clarifications observed for the NCP groups, simply because of the nature of the interactive treatment activities. For example, while playing Category Card Games it was common for a child to mislabel a card’s category and for the clinician to provide the correct category (or to provide clues as to the correct category). In this scenario the clinician either corrects or expands the child’s utterance semantically, and a linguistic cue would be counted. Tallying linguistic cuing during the EL sessions provided a measure of how much language-based stimulation these participants were receiving, in addition to the linguistic stimulation from the computer and scripted portions of the interactive activities.

As expected, there was very little linguistic cuing during NCP sessions ($M = 0.2$). During the EL sessions, linguistic cuing was more common ($M = 16.4$). The
difference between NCP and EL sessions was significant \( t (31) = 12.4, p < 0.001, d = 1.7 \).

**Analyses**

As indicated in Chapter 4, three types of change (individual, absolute, and relative) were examined in each of three areas (nonlinguistic cognitive processing, English language, and Spanish language). Each type of change required a different analysis. Analyses were conducted nonparametrically due to the relatively small numbers of participants in each group and the lack of evidence that the outcome variables are normally distributed.

**Individual change.**

The purpose of the individual change analyses was to determine whether each child made real or reliable change on each task. To meet this goal, a criterion for real change was set for each task. Each child was then assigned one of three scores: a + if his or her change from pre-testing to post-testing met or exceeded the criterion in a positive direction; a – if his or her change from pre-to post-testing met or exceeded the criterion in a negative direction; and a 0 if neither of the other two conditions applied. In other words, a + score indicated that the participant’s task score had truly changed rather than randomly fluctuated from pre- to post-testing.

The criteria for change were based upon the standard error of measurement (SEM) for the task, derived from peer data. For standardized language tests, published confidence intervals provided the needed criteria. Additional peer data were obtained from Kohnert, Windsor, and Pham (2009); this database contains the performance of
over 200 children on the same choice visual detection, auditory serial memory, and Spanish and English NWR tasks as were used in the present study. More specifically, a total of 74 Spanish-English bilingual children between 6 and 10 years old participated in the Kohnert, Windsor, and Pham (2009) study. Nineteen of these children had PLI and 55 had typical language development. The standard deviation for each 1-year band of Spanish-English bilingual participants in the study (i.e., 6-year-olds, 7-year-olds, 8-year-olds, and 9-year-olds) was calculated to derive an age-specific estimate of task variability. When data from Kohnert, Windsor, and Pham (2009) was not available, the overall standard deviation from the current study was used. Then an estimate of reliability for each task was calculated in order to derive the SEMs. The specific considerations in deriving criteria for each task are described in detail below.

Language assessments.

For all standardized language assessments, only standard scores (rather than raw scores) were analyzed. The criterion for change was set at the 90% confidence interval. Thus, if the child’s retest score exceeded the 90% confidence interval, the change was considered real and the child was assigned a + for that test. As discussed above, the Spanish vocabulary measures (EOW-S and ROW-S) were inadvertently administered inconsistently for some participants. An individual change score was calculated only for participants who had comparable scores from pre- to post-testing (i.e., completed the test in Spanish at both pre- and post-testing or completed the test bilingually at both pre- and post-testing). In addition, two children in the delayed control group turned 9 years old between the pre- and post-testing. Examiners followed the CELF-4E and CELF-4S
procedures, which call for administering the *Word Structure* subtest to 8 year old children and the *Word Classes* subtest to 9-year-old children. Thus this subtest was inconsistent from pre- to post-testing for two participants and therefore was not analyzed for them.

Published confidence intervals were used because they provide a conservative, conventional measure of change for the individual analyses. However, one drawback to using published confidence intervals is that a few of the children in this study performed near floor levels on some of the English standardized tests, which are not normed for bilingual children. In other words, despite some gain in raw score from pre- to post-testing, the standardized scores were not sensitive to these gains and these children maintained the lowest possible standard score.

For the two nonstandardized assessments, peer data from Kohnert, Windsor, and Pham (2009) were used to derive the SEM and resulting criteria. The English and Spanish NWR tasks were identical to those used in the peer database (see also Windsor, Kohnert, Lobitz, & Pham, 2010), and the standard deviation by age band (as discussed above) was calculated from the database. Split-half reliability was calculated from data in the current study, and a 90% confidence interval was created.

The RAN task used in the Kohnert, Windsor, and Pham (2009) study was substantially different from the RAN used in the current study. However, the RAN task here was very similar to the typical administration of the RAN tasks in the CELF-4E and CELF-4S. Therefore the overall standard deviation of RAN scores in the current study, along with split-half reliability estimates published in the CELF-4E and CELF-
4S manuals, was used to create the SEMs and resulting 90% confidence intervals for the Spanish and English RAN tasks.

**NCP assessments.**

The choice visual detection task was again identical to the one in Kohnert, Windsor, and Pham (2009), and the procedures for calculating the 90% confidence interval follow those for NWR exactly. The auditory serial memory task departed from this procedure. As discussed above, the achievement level variable was created to adequately capture change across the wide variety of performance seen in this study. A criterion of plus or minus one achievement level was set as the standard for real change.

The sustained selective attention task was created for the current study. Therefore no peer data was available for this specific task. The standard deviation of the corrected $d'$ value across all age groups in the current study was calculated, along with the split-half reliability, to create the 90% confidence interval.

When this criterion was applied to the results, however, there was a near-perfect negative correlation between pre-test score and real change score. In other words, children who initially scored well on the test were not able to make enough change in their accuracy score to meet the criterion for change. This situation resulted from a substantial ceiling effect for task performance in this sample. Therefore, a second criterion based on RT was set up for children who demonstrated accuracy near ceiling at the pre-test. No participant who scored above 390 (of 400) items correct at the pre-test met the criterion for change. Therefore, participants who scored at least 390 and maintained high accuracy (i.e., they did not meet the change criterion for $d'$ in a
negative direction) were subject to the RT criterion. The RT criterion for positive change was a decrease in mean RT of at least 15% of the mean pre-test RT.

**Absolute change.**

To analyze absolute change, the Fisher matched-pairs exact test was conducted for each measure for each of the three groups. The Fisher matched-pairs exact test is a nonparametric analog of the matched-pairs t-test; it provides the probability that the pre-to post-test change on a given measure within a group is due to chance. The test is at least as efficient as the t-test, and possibly more so for small samples, but the calculations required increase exponentially as sample size increases. To calculate the probability that change is due to chance, all possible values of the test statistic (in this case, the mean pre- to post-test change score) are calculated, assuming that the sign on each observed change score is equally likely to be either positive or negative when change is driven by chance. Then the probability of the observed mean change score is the number of possible mean scores equal to or greater than the observed score, divided by the number of possible change scores. In this study, the Fisher test was executed using NPStat 3.8 software (May, Hunter, & Masson, 1993). One-tailed analyses were performed as the predicted direction of change is universally positive for the treatment groups. In addition, the standardized mean difference between pre- and post-test scores was calculated as an effect size for each pre- to post-comparison (Cohen, 1988).³

² The null hypothesis for nonparametric tests such as the Fisher matched-pairs exact test is equality of distributions rather than equality of a single location parameter such as group mean. The use of effect sizes such as Cohen’s $d$ with this test thus makes the additional assumptions that dispersion and
For each measure, the primary dependent variable (listed in Tables 3 and 4) was used in the analysis. Despite the ceiling effect described in the section above on Individual Change, corrected $d'$ remained the best overall dependent variable for sustained selective attention. Standard scores were used for all standardized language tests to make this analysis relatively conservative in detecting absolute change. Finally, only the three subtests common across both CELF-4 age groups (i.e., *Recalling Sentences, Formulated Sentences*, and *Concepts and Following Directions*) were analyzed. The scores from *Word Structure* or *Word Classes* were omitted because they were inconsistent across children in a group.

**Relative change.**

To measure relative change, a three-way comparison across groups was conducted for each measure. Pre-test scores and age in months were included as covariates to remove the potentially confounding effects of these variables. The comparison was conducted using the nonparametric version of ANCOVA, an Aligned-Rank procedure (Hettmansperger, 1984). In this procedure a regression model is created for each measure, with the pre-test scores and age as predictors and the change scores as the dependent variable. The unstandardized residuals are obtained from this model; the residuals are free from the effects of the pre-test scores and age. The residuals are then ranked and used as the dependent variable in a new regression model, with the treatment groups as predictors. The resulting sums of squares are used to

---

shape are the same for the distributions being compared. In this case, the cost of such assumptions was judged to be less than the benefit of reporting effect sizes.
calculate a test statistic that follows a chi-square distribution. IBM SPSS Statistics version 19.0 was used to conduct the regressions and ranking in the Aligned-Rank procedures (SPSS Inc., 2010), with hand calculation of the test statistic.

In order to make the relative change analyses more sensitive to subtle differences between treatment groups, raw scores were used in place of standard scores for all standardized language measures. As with the absolute change analyses, the relative change analyses were conducted for only the three CELF-4 subtests consistent across all ages.

**Treatment satisfaction outcomes.**

Finally, to explore the role of participant and family treatment satisfaction in outcomes, the parent and child satisfaction surveys were converted into treatment satisfaction scores. For each survey, responses were quantified by assigning a numerical scale to the available responses, with higher numbers indicating greater satisfaction. For the parent surveys, “Agree Strongly” was coded as 5 and “Disagree Strongly” was coded as 1; for the child surveys, “Yes” was coded as 3, “Maybe” as 2, and “No” as 1. Scores for each question were then summed across the survey to create a single overall score for each participant and parent survey.

The lack of variation in the child satisfaction survey results precluded completion of the planned statistical analysis of this data. Out of 18 surveys and 72 total questions administered, only two responses were not “yes”. These responses indicated children overwhelmingly enjoyed the treatment activities, liked the clinician administering the treatment, and perceived some benefits (“listening and talking better”).
as a result of participating in either the EL or NCP treatment. Given the homogeneity of responses, no statistical analysis of the child satisfaction surveys was completed.

Responses on the parent surveys were slightly less homogeneous. A total of 13 survey responses were collected; the remaining five parents of participants in the NCP and EL groups could not be reached by phone at the conclusion of treatment. When responses for each of the eight questions were summed, the resulting total survey scores ranged from 28 to 40 (out of 40 possible points). The total survey scores for the NCP and EL groups were compared using the two-sample Fisher exact test. Like the Fisher matched-pairs exact test, the two-sample Fisher exact test uses permutations of the observed scores to calculate the probability that change is due to chance. The test was conducted using NPSTAT (May et al., 1993).

Finally, the relationship between treatment satisfaction scores and total change in NCP, English, and Spanish skills was calculated using nonparametric Spearman’s rank correlations. The nonparametric correlations were used due to the clear deviation from normality in the distribution of the total survey scores. Total change scores in each of the three outcome areas (NCP, English, and Spanish) were calculated by summing the ‘+’s and ‘−’s that were generated for the individual analyses (see Figure 2 for further details). The sum within each category became a gross index of change in that area. These NCP change, English change, and Spanish change indices were then used in the correlation analysis with total satisfaction survey score.
Post-hoc analyses.

Two factors emerged during the study as potentially important influences on treatment outcomes: age and treatment attendance. These factors were explored in post-hoc analyses. Parametric statistics were used for these explorations because the treatment groups were combined, creating a larger group, and because the creation of gross change indices for NCP, English language, and Spanish language skills created more normal dependent variables.

The first factor, age, became potentially important when a correlation between age and pre-test performance was discovered (Payesteh, Rentmeester, & Kohnert, 2011). The correlation holds even for standardized test scores, which should account for age. Correlations between pre-test scores and age were stronger for English than for Spanish. The pattern may reflect, in part, the “mismatch” between the bilingual participants in this study and the monolingual English-speaking normative samples for the English standardized tests that were used. In addition, the positive correlations between age and English standardized scores are consistent with literature indicating that TD bilingual children in the US experience rapid growth in English during the school years (e.g., Kohnert & Bates, 2002; Pham, 2011; see Chapter 7 for further discussion).

Because of the relationship between pre-test score and age, the relationship between pre- to post-test change and age was explored. Pearson correlations between each of the three change indices (NCP, English, and Spanish, described above) and age in months were calculated for the children in the two active treatment groups.
Finally, an exploratory analysis of the relationship between treatment change and treatment session attendance was conducted due to the relatively large variability in session attendance amongst the members of the active treatment groups. The minimum session attendance criterion for inclusion in a treatment group was set at 16; using a higher number would result in unacceptable attrition levels. However, three children in Cycle 1, who had perfect attendance, participated in 24 sessions. To determine whether the number of sessions attended above the critical minimum of 16 was related to treatment gains, a regression analysis was performed for each of the three overall indices of change (NCP, English, and Spanish), with number of sessions attended as the sole predictor.
Chapter 6: Results

Individual Change

The purpose of the individual change analyses was to examine within-group variability on each dependent variable, in the event that not all individuals in the group follow the same patterns. Full results of the individual change analyses are presented in Tables 8a, 8b, and 8c. In addition, a graphic summary of the individual change analyses is presented in Figure 2. This graph displays the percentage improvement for each individual in English and in Spanish. Percentage improvement for each language was calculated by summing the +’s and –‘s within a category, counting + as 1 and – as -1, and dividing by the total number of tasks in the category. In addition to the tables and graphic summary, a descriptive summary of the individual change data is presented here.

NCP assessments.

The Choice Visual Detection task served as the primary measure of processing speed for nonlinguistic stimuli. On the Choice Visual Detection task, 5 of 10 individuals in the NCP treatment group achieved the criterion for significant change. Within the EL treatment group, 3 of 8 children achieved this criterion. Two of six children in the control group achieved significant change. There were no children in any group who exceeded the criterion for negative change; that is, no child became significantly slower from pre- to post-testing on this task.

The Auditory Serial Memory task was included to measure short-term or working memory for nonlinguistic stimuli. The NCP treatment was not designed to
target this cognitive skill. In the NCP treatment group, 2 of 10 children made positive change and 1 of 10 made negative change; in the EL treatment group, 6 of 8 children made positive change and 1 of 8 made negative change; and in the control group 3 of 6 made positive change and 1 of 6 made negative change.

The Sustained Selective Attention task was designed to measure participants’ ability to focus and detect target stimuli over an extended period of time. Both positive and negative change occurred within the NCP group on this task; 5 of 10 children achieved the positive change criterion, but 3 of 10 exceeded the criterion for negative change. Within the EL treatment group, change on the Attention task was almost universally positive: 6 of 8 children made positive change and no child met the criterion for negative change. Finally, the control group demonstrated unexpected change on this task. All children in the group reached either the positive or negative change criterion, with 5 children making positive change and 1 child making negative change.

Overall, the individual analyses for NCP measures do not support a substantial advantage for NCP treatment over EL treatment or delayed treatment. On each task, some children in each of the three groups improved and some did not. Thus, there was substantial within-group variability. In the area of memory, which was not specifically targeted by the NCP treatment, there is a notable trend towards improvement in the EL group that is not present in the NCP group.

Also, in general the delayed treatment control group demonstrated more significant change than anticipated. The control group did contain a slightly greater
proportion of instances of negative change; however, the overall trend within the delayed treatment group was for positive change on the NCP measures.

Finally, all children made positive change on at least one of the three NCP tasks. Only two children made positive change on all three measures. These figures indicate substantial variability within individuals in performance across the three tasks.

**English language assessments.**

The trends toward variability within individuals, groups, and tasks continued in the individual change analyses for the English Language assessments. Within the NCP group, no more than 4 of 10 children achieved the positive change criterion on any single measure, and no child achieved the positive change criterion on more than 6 of 9 tasks. However, all children in the group made positive change on at least one task and at least one child made positive change on every task. Thus, the overall group trend is for improvement on English language measures, despite the relatively conservative standards used in these analyses.

There are a few instances of negative change occurring on English measures within the NCP group. Seven of ten children in the group showed negative change on at least one English task. In particular, the performance of three children declined on the Concepts and Following Directions subtest, and the performance of three children also declined on the RAN-E (though not the same group of three children). While it is possible that the NCP treatment did not promote growth in these areas, individual variability again interferes with that pattern, as at least two children in the group achieved the positive change criterion on both measures.
Within the EL treatment group, similar patterns hold. The overall trend is for positive change, and every child in the group makes positive change on at least one task. Positive change appears on every task, with the exception of the Word Classes subtest of the CELF-4E (which only two children completed). For two tasks, six of eight children in the group achieved the positive change criterion: the Recalling Sentences subtest of the CELF-4E and the EOW-E. Only two children in the EL group demonstrated negative change on an English language task.

In the delayed treatment group, the patterns shift slightly. Individual variability continues, and both positive and negative changes are seen in this group despite the lack of treatment. However, the distribution of significant positive change is somewhat less pervasive than in the two active treatment groups. Two of the six children did not demonstrate positive change on any of the English tasks. In addition, no children in the group achieve the positive change criterion on three of the tasks: the Word Structure subtest from the CELF-4E, the Formulated Sentences subtest from the CELF-4E, and the NWR-E task.

Nonetheless, there is substantial positive change at the individual level within the delayed treatment group. On two of the tasks (RAN-E and EOW-E) at least three of the six children, or 50% of the group, reached the positive change criterion. One child demonstrated positive change on four of the eight measures he completed.

**Spanish language assessments.**

The individual change results for the Spanish language assessments again present a mixture of positive and negative change, with positive change predominating.
In the NCP group, nine of ten children achieved the positive change criterion on at least one Spanish task. No child achieved it for more than five Spanish tasks. Once again, for all measures at least one child made positive change. The greatest concentration of positive change occurred on the RAN-S, with four of the seven children who successfully completed the task making positive change, and on the EOW-S, with five of the nine who had consistent pre- to post- test administration making positive change. In addition, while only four children completed the CELF-4S Word Classes subtest due to age, three of them achieved positive change on it.

Six of the ten children met the negative change criterion on at least one Spanish language task. No child demonstrated negative change on more than two of the measures, and no more than two children made negative change on any single task. Thus, negative change was present in the NCP group but was distributed across individuals and measures with relatively low frequency. Negative change and positive change were also comparable in both English and Spanish in this group, with 25 instances of positive change in English versus 28 in Spanish, and 10 instances of negative change in English versus 9 in Spanish.

Like the NCP group, the EL group contained one individual who did not achieve positive change on any Spanish language task. Also like the NCP group, for each measure the EL group contained at least one child that achieved positive change. Overall, the distribution of positive change in the group was roughly comparable to the group’s positive change in English (25 instances of positive change in Spanish versus 31 in English, with 5 missing measures in Spanish). The measures with the greatest
proportion of positive change were the CELF-4S Concepts and Following Directions and Word Classes subtests, and the NWR-S task.

However, the occurrence of negative change in the EL group jumped dramatically in Spanish in comparison to English. In Spanish, 17 instances of negative change were seen across the group, with only 4 instances on the English measures. In particular, two children appeared to demonstrate some language loss with the EL treatment: participant 13 met the negative change criterion on two measures and the positive change criterion on none, and participant 24 met the negative change criterion on six of the eight Spanish language measures he completed.

For the delayed treatment control group, the individual change analyses for the Spanish language measures showed substantial volatility. Despite the existence of 6 missing measures in Spanish, the occurrence of both positive change and negative change within the group increased from the English to the Spanish analyses. In English, there were 12 instances of positive change in the group; in Spanish, there were 17. In English, there were 4 instances of negative change in the group, and in Spanish there were 15. Like the EL group, the delayed treatment control group had one child who demonstrated negative change across nearly all Spanish measures (six of eight). In contrast, another child in the group made positive changes on six of the Spanish language measures. Thus, the Spanish language individual change analyses for this group reveal variability in language growth between individuals, even in the absence of study-based treatment.
Overall, the individual change analyses indicate the presence of notable individual differences in language and cognitive skill change over time. While every participating individual demonstrated growth from pre- to post-testing in at least one area, generalizations are difficult to create because patterns of change are not consistent within a group for any of the three outcome areas. The presence of this individual variability provides an important context for the group level analyses of absolute and relative change.

**Absolute Change**

The absolute change analyses used the Fisher matched-pairs exact test to determine whether the change from pre- to post-testing for each group on each measure was significant. Full results for the Fisher tests are reported in Table 9a for the NCP treatment group, in Table 9b for the EL treatment group, and in Table 9c for the delayed treatment control group.

**NCP assessments.**

For the NCP treatment group, the Fisher test reached significance only for the choice visual detection task (\(d = 0.86, p = 0.04\)). The tests for auditory serial memory (\(d = 0.20, p = 0.38\)) and sustained selective attention (\(d = 0.28, p = 0.10\)) did not reach statistical significance, though mean group change was in the expected direction for both tasks.

For the EL treatment group, the Fisher tests were significant for all three NCP assessments: auditory serial memory (\(d = 0.63, p = 0.047\)), choice visual detection (\(d = 1.01, p = 0.02\)), and sustained selective attention (\(d = 0.90, p = 0.008\)). Thus the EL
treatment group demonstrated significant improvement from pre- to post-testing in the three NCP areas of interest, memory, processing speed, and attention.

No significant change was observed in the delayed treatment control group on any of the three variables (auditory serial memory, \( d = 0.33, p = 0.25 \); choice visual detection, \( d = 0.60, p = 0.25 \); sustained selective attention, \( d = 0.39, p = 0.06 \)). However, mean group change was in the positive direction on all three tasks, and the decreased power in this group (due to the smaller group size) may have contributed to the lack of statistical significance.

**English language assessments.**

The NCP treatment group demonstrated significant pre- to post- treatment change on two English language measures: the *Formulated Sentences* subtest from the CELF-4E \( (d = 0.54, p = 0.03) \) and the Core Language composite score from the CELF-4E \( (d = 0.41, p = 0.008) \). Mean group pre- to post- change was in the predicted positive direction on all English tasks except NWR, in which total PPC dropped slightly, and the CELF-4E *Concepts and Following Directions* subtest, in which there was no pre- to post- group change.

For the EL group, absolute change was significant for three English language measures. The mean group change on the EOW-E was eight standard score points \( (d = 0.85) \) and was significant at \( p = 0.008 \). In addition, the *Recalling Sentences* subtest of the CELF-4E showed significant change \( (d = 0.72, p = 0.02) \), as did the overall Core Language composite \( (d = 0.45, p = 0.02) \). Finally, mean group change on the RAN-E
neared significance ($d = 0.69, p = 0.06$). All English language change scores for the EL treatment group reflected positive pre- to post- treatment change.

For the delayed treatment group, none of the absolute change analyses for English language measures reached significance. Group pre- to post- change was positive for all measures, except two CELF-4E subtests: *Concepts and Following Directions* and *Recalling Sentences*. The group showed no change on these subtests.

**Spanish language assessments.**

None of the Fisher tests for the Spanish language measures reached statistical significance for the NCP treatment group. The results did approach significance for both NWR-S ($d = 0.42, p = 0.07$) and the CELF-4S Core Language composite score ($d = 0.17, p = 0.07$). Change was in the expected positive direction for all measures.

Results of the Fisher tests for Spanish language measures for the EL treatment group followed a similar pattern. None of the tests reached statistical significance at the $p < 0.05$ level. The test for the CELF-4S Core Language composite score approached significance ($d = 0.32, p = 0.07$). The EL treatment group made positive change on all Spanish measures except the ROW-S, for which the mean group score dropped by five standard points, and the CELF-4S *Formulated Sentences* subtest, for which the mean group score dropped by less than one scaled score point.

Finally, none of the Fisher tests for the Spanish language measures for the delayed treatment group reached significance. The results for the RAN-S were nearly significant ($d = 0.92, p = 0.06$) but based on a very small group of participants; all four of the children in the delayed treatment group who successfully completed the RAN-S
improved upon it from pre- to post-testing. There were four Spanish language measures for which the delayed treatment group’s average change was negative: the CELF-4S Formulated Sentences subtest, the CELF-4S Core Language composite, the ROW-S, and the EOW-S.

**Relative Change**

The relative change analyses compared all three groups on each outcome measure, using the pre-test scores and ages as covariates. The raw scores from standardized testing were used in these analyses to increase sensitivity to change between groups. A summary of the results of these tests appears in Table 10.

**NCP assessments.**

The Aligned-Rank procedure did not yield significant results for any of the three NCP assessments. For the choice visual detection task, the amount of variance accounted for by group membership was quite small (0.5%; \( p = 0.95 \)). The amount of variance in change on the sustained selective attention task that was accounted for by group membership was slightly larger (13.3%; \( p = 0.25 \)). Group membership accounted for 10.0% of variance in auditory serial memory task outcomes (\( p = 0.55 \)). On all three tasks, mean rank was the highest for the EL treatment group, indicating a relative advantage for this group when age and pre-test score are covaried out of change scores.

**English language assessments.**

For English language assessments, two of the Aligned-Rank procedures were significant. On the CELF-4E Recalling Sentences subtest, group membership accounted for 36.1% of the variance in change scores (\( p = 0.02 \)). Post-hoc group
comparisons indicated that change in the EL group exceeded change in the NCP group on this measure, with no other group comparison reaching significance. On the CELF-4E Core Language composite, group membership accounted for 30.9% of the variance in change scores, which was a significant result at $p = 0.02$. Post-hoc group comparisons indicated that the EL group outperformed both the NCP and delayed treatment groups on this measure.

For the remainder of the English language measures, relative change was not significant for any group, though the relatively small numbers in the analysis may affect the power to detect it. The percentage of variance accounted for by group membership exceeded 10% for three other measures: CELF-4E Concepts and Following Directions (16.3%, $p = 0.15$), CELF-4E Formulated Sentences (14.4%, $p = 0.19$), and NWR-E (22.9%, $p = 0.09$). For the Formulated Sentences subtest, the mean group ranks suggest a trend towards better performance in the NCP group; for Concepts and Following Directions and NWR-E, the mean group ranks suggest better performance in the EL group.

**Spanish language assessments.**

The relative change analyses indicated no significant differences between groups on any of the Spanish language measures. The percentage of variance accounted for by group membership exceeded 10% on three measures. For the EOW-S, group membership accounted for 11.5% of variance ($p = 0.30$), with the group ranks suggesting comparable change in the EL and NCP groups and decreased change in the delayed treatment group. For the CELF-4S Formulated Sentence subtest, results appear
to mirror the English relative change results: group membership accounted for 16.9% of variance \( (p = 0.14) \) and group ranks reflect greater change in the NCP group, with the least change in the delayed treatment group.

Finally, group membership accounted for 12.1% of the variance in CELF-4S Core Language composite score \( (p = 0.26) \). The mean group ranks are quite similar to those for the EOW-S, suggesting a trend towards greater change in the two active treatment groups than in the delayed treatment group. In fact, the delayed treatment group shows the lowest mean rank (indicating relatively less change) on all measures except the ROW-S and the CELF-4S Recalling Sentences subtest.

**Treatment satisfaction**

A group comparison of the child satisfaction survey results was not performed due to the consistency of survey responses. The mean total survey score was 11.83 out of 12. Though this measure was clearly not sensitive to between-group differences in child satisfaction with the treatment program, it did indicate overall positive associations with the experiences for children in both active treatment groups.

For the parent satisfaction surveys, survey responses also tended to be positive, but slightly more variable than the child surveys. A comparison of the NCP and EL groups was conducted, using the two-sample Fisher exact test. The mean total survey score for the 6 families in the NCP group who could be reached was 37.0 out of 40 (SD = 2.8). The mean for the 7 families in the EL group who could be reached was 35.9 out of 40 (SD = 4.3). The difference between the two groups was not significant \( (p = 0.62) \); using a two-tailed analysis, there are 1069 permutations of the scores that would
generate a test statistic more extreme than the observed results, out of 1716 total possible permutations.

The Spearman correlation between total satisfaction survey score and NCP change was not significant ($\rho = 0.13, p = 0.68$). Neither was the correlation between survey score and Spanish change ($\rho = 0.30, p = 0.32$). However, the total English change index was significantly correlated with total survey score ($\rho = 0.63, p = 0.02$).

**Post-hoc Analyses**

**Age**

The correlations between participant age in months, NCP change, English change, and Spanish change were calculated. There was a significant correlation between age and Spanish change ($r = 0.60, p = 0.009$) and between age and NCP change ($r = -0.62, p = 0.006$). The relationship between age and English change was also negative ($r = -0.35$) but did not reach significance ($p = 0.17$). Thus, the younger children in the study tended to improve substantially more on the NCP tasks than the older ones did, and to improve somewhat more than older children on English assessments. The older children improved more on the Spanish measures, however.

**Attendance.**

The number of sessions attended did not vary between the two treatment groups. Participants in the NCP group attended an average of 20.5 sessions (SD = 2.2) and participants in the EL group attended an average of 20.8 sessions (SD = 2.9). When the two treatment groups were combined, attendance did not predict NCP gains ($R^2 = 0.005, p = 0.77$) or English language gains ($R^2 = 0.002, p = 0.87$). Attendance predicted
more variance in Spanish language gains but results did not reach statistical significance
($R^2 = 0.16, p = 0.10$).
Chapter 7: Discussion

Revisiting Research Questions

The purpose of this study was to investigate the effects of two treatments for bilingual children with PLI. There were three groups of research questions outlined in Chapter 4, which will be considered in turn here.

Effects of treatment on NCP skills.

The first group of research questions related to the effects of treatment on NCP skills. It was hypothesized that treating processing speed and sustained selective attention through the NCP treatment would lead to gains in these skills. Auditory working memory was also measured as a third cognitive skill that could change after treatment, though it was not directly targeted by the NCP treatment.

There was some evidence for change in NCP skills in the NCP treatment group according to the individual and absolute change analyses, but not according to the relative change analyses. The evidence was strongest for change in speed of processing, as measured by the choice visual detection task, and weakest for change in auditory working memory, as measured by the auditory serial memory task. The lack of dramatic change in working memory for this group is consistent with the original hypotheses: the NCP treatment was not designed to affect working memory. Memory was a substantial component of only one game in the treatment protocol (Simon Trickster). Instead, the treatment focused on processing speed and attention, which did appear to be more strongly affected. In particular, gains on the choice visual detection task reached significance and reflected a “large” effect size (Cohen, 1988). However,
the absence of a relative advantage for this group in processing speed and attention gains was inconsistent with the original hypotheses.

For the EL treatment group, evidence for NCP skill change was also apparent at the individual and absolute change levels. The EL treatment group demonstrated relatively greater change than the other two groups on all three NCP tasks, as indicated by the higher mean rank for the group in the Aligned-Rank procedure, but this advantage did not reach statistical significance. The strong performance of the EL group on the absolute change analyses, as well as the trend towards a relative advantage for this group on all three tasks, suggests that all three cognitive skills of interest appeared to improve through the EL treatment.

This outcome was not intended to be a focus of the study. Yet it is entirely consistent with the suggestions discussed in Chapter 2 that intensive language treatment may modify information processing skills such as attention (e.g., Gillam et al., 2008). The EL treatment group was asked to focus on a treatment activity for the duration of their after-school program just as the NCP group was. This focus could lead to improvements in sustained attention. The inclusion of games that focus on following directions and processing auditory information (such as Calling All Engines) present an opportunity for growth in auditory memory. Improvements in processing speed are perhaps the most difficult to explain for this group, as the activities in the EL treatment did not emphasize speed of response. Overall, the EL group’s gains in NCP skills suggest that NCP skills may be an inherent aspect of many language-based treatment
activities; we removed language from NCP treatment here, but it may not be possible to remove NCP skills from language treatment.

The delayed treatment group demonstrated unexpected positive change on all three NCP tasks. Though the absolute change analyses did not reach significance for any of the three tasks, there were group trends towards task improvement as well as several individuals who met the positive change criterion for each task. This trend towards positive change for the delayed treatment group impacted the results of the relative change analyses; the positive change in the two active treatment groups was not statistically different from the delayed treatment group, as was predicted by the initial hypotheses.

A combination of factors may have lead to the delayed treatment group’s improvement on NCP tasks. It is possible that experiences outside the experimental treatment were leading to improvements in processing speed, working memory, and sustained selective attention. All children in the study were attending school and receiving school-based speech-language services, which may improve the NCP skills assessed. Age and maturation may also contribute to growth on these non-standardized tasks. Though the time between pre- and post-testing was relatively short in this study, age is known to be a robust contributor to performance on the choice visual detection and auditory serial memory tasks (Kohnert & Windsor, 2004; Kohnert, Windsor, & Pham, 2009).

Task-related factors may have also influenced performance on the NCP tasks. The general trend towards improvement across all three groups and all three tasks
suggests that some practice effects may occur on these tasks. The examiners administering the pre- and post-testing reported that at post-testing children tended to recall having completed the computerized tasks before and that the children required substantially less training on the tasks at post-testing. These reports lend anecdotal support to the possibility of practice effects. Although two of the NCP tasks used here (choice visual detection and auditory serial memory) are sensitive assessments for PLI in the population of interest (Kohnert & Windsor, 2004; Kohnert, Windsor, & Pham, 2009), the existing databases are cross-sectional rather than longitudinal. They do not provide insight into typical task performance on repeated administrations. The third task, sustained selective attention, was created for this study and therefore does not yet have substantial cross-sectional or longitudinal data associated with it. Thus, it is possible that some component of improvement on the three NCP tasks was due to repeated testing rather than true improvement.

However, the lack of consistent improvement across all or even most individuals is apparent in Table 8c. Many children did not improve performance on these tasks, or even performed more poorly at post-testing than at pre-testing. In addition, the five children who completed each task three times (by virtue of participation in the delayed treatment control group and then in an active treatment group) do not present a clear picture of improvement with each subsequent test cycle. These data suggest that any potential practice effects are not universal and contribute just one consideration to the interpretation of performance at post-testing.
An additional task-related factor that may complicate interpretation of the NCP assessment results is the sensitivity of the dependent variables. The substantial range of task performance across all study participants made the creation of sensitive tasks and dependent variables extremely challenging. This was particularly problematic for the sustained selective attention task. Despite the adjustments made to task difficulty following pilot testing (described in Chapter 3), ceiling effects persisted for the planned dependent variable, $d'$. For example, 12 of 29 pre-test accuracy scores for the task fell at or above 390 items correct out of 400 possible items. This left inadequate opportunities for some children to demonstrate improvement; at the same time, two children were unable to comprehend and complete the task at pre-testing, illustrating the range of performance across the group. In the individual level analyses, these differences could be accommodated somewhat by using different criteria for change depending upon pre-test score. However, a single dependent variable was necessary for group analyses, and this dependent variable may not have been optimally sensitive to change for all participants. To a lesser degree, this condition may apply also to choice visual detection -- on which a few children made accuracy gains that were not considered in the analyses -- and to auditory serial memory. The achievement level variable was created to improve auditory serial memory task sensitivity across ability levels, but it is possible the variable could be improved with further refinement.

In summary, the results of the NCP assessments present a somewhat complicated picture. The delayed treatment control group was the only group not to achieve statistically significant change on any of the three tasks in the absolute change
analyses, but the group did demonstrate positive change on all three tasks. The two active treatment groups did make significant gains on the NCP tasks (on all three tasks for the EL group and on the choice visual detection task for the NCP group), but they did not statistically exceed the control group’s changes. The lack of a significant difference between the two active treatment groups and the control group on NCP measures is likely due to the small sample size. Finally, the individual change analyses clearly demonstrate substantial variability in task gains across children, even within each group. On the whole the results are consistent with previous work indicating that attention, memory, and processing speed in populations of children with disabilities can be modified through treatment (e.g., Cosper et al., 2009; Kerns et al., 1999; van’t Hooft et al., 2007). However, the pattern of results may reflect multiple influences, including the potential effects of school and school-based treatment, of repeated testing and maturation, of task sensitivity across a wide range of abilities, and of individual responsiveness to treatment.

**Effects of treatment on English language skills.**

The second group of research questions related to the individual, absolute, and relative effects of treatment on the English language skills of participants. It was hypothesized that NCP treatment would lead to change via the influence of improved NCP skills and that EL treatment would lead to change by targeting the English language skills directly. For the delayed treatment control group, the effects of school and school-based language treatment have the potential to affect English language
skills; it was hypothesized that the intensity of the experimental treatments would be sufficient to create differential effects across groups.

The results presented here provide evidence of some change in English language skills within the NCP treatment group. Change in the CELF-4E Core Language score, which compiles change across the four subtests, provides the most compelling evidence of improvement. The NCP group improved CELF-4E Core Language scores from pre-to post-testing at the $p < 0.01$ level, with the group gaining nearly one-half a standard deviation from pre- to post-testing ($d = 0.41$). In addition, three children exceeded the 90% confidence interval in the positive direction and no individual child exceeded it in the negative direction. Significant pre- to post-test change was also seen on the Formulated Sentences subtest, with trends towards significance on the English vocabulary measures (ROW-E and EOW-E); effect sizes for Formulated Sentences ($d = 0.54$) and the ROW-E ($d = 0.70$) could both be considered “medium-sized” (Cohen, 1988). Finally, the individual analyses suggest progress on the WS/WC combined subtest as three children earned a ‘+’ and none earned a ‘–’. This subtest was not included in absolute or relative change analyses because the administration depends on age.

The pattern of progress in English language skills for the NCP group is quite similar to the gains seen in the NCP treatment for monolingual children in the SSED pilot study (Ebert & Kohnert, 2009; summarized in Chapter 3). In the pilot study, standardized testing gains were driven by improvements in the Formulated Sentences and Word Structure subtests from the CELF-4E, and in the EOW-E. One potential
interpretation of that pattern is that the treatment facilitated the ability to retrieve both words and morphosyntactic knowledge (Ebert & Kohnert, 2009).

Similarly, the English gains for the NCP group here are seen on tasks with an emphasis on linguistic content and structure, rather than on linguistic memory. The NCP group’s poorest performance comes on the tasks with the greatest memory requirements: NWR-E, CELF-4E Recalling Sentences, and CELF-4E Concepts & Following Directions. The lack of gains on memory-intensive tasks is again consistent with our previous work using NCP treatment (Ebert & Kohnert, 2009).

In the relative analyses, the English language gains seen in the NCP group did not exceed the gains made in other groups when the effects of age and pre-test score were considered. The NCP group attained the highest relative rank for the ROW-E, the CELF-4E Formulated Sentences subtest, and the RAN-E, but none of these differences reach statistical significance. In addition, the English language gains made by the EL treatment group exceed the NCP group’s gains on the CELF-4E Core Language score. Thus, while the NCP group made gains in English language skills, there was a relative advantage for overall gains in English language skills for the EL treatment group.

In the EL treatment group, there is substantial evidence of English language gains, across the individual, absolute, and relative levels of analysis. The group as a whole makes positive change on all English measures, with statistically significant change occurring on the EOW-E, the CELF-4E Recalling Sentences subtest, and the CELF-4E Core Language composite. The group demonstrates the highest mean rank on
all English measures except the three previously mentioned which show an advantage for the NCP group.

The pattern of English gains in the EL group is distinct from the NCP group. The gains are particularly striking for the *Recalling Sentences* subtest and the EOW-E; on each of these tasks, 6 of the 8 children in the group exceeded the 90% confidence interval in the positive direction; no children made negative change on either task; and effect sizes for both tasks are large ($d = 0.72$ and $d = 0.85$). The group also makes positive, though not quite significant, gains on the CELF-4E *Concepts & Following Directions* subtest and the NWR-E. A common element across *Recalling Sentences, Concepts & Following Directions*, and the NWR-E is the memory requirement. One interpretation of the EL treatment group’s gains is that memory for language improved. In addition, vocabulary appears to have improved, as evidenced by the gains on EOW-E.

Thus, this traditional English language treatment appears to benefit bilingual children with PLI. Given that the treatment was designed to improve English vocabulary, morphosyntax, and linguistic auditory memory, this result is perhaps not surprising. It is nonetheless encouraging to have evidence that widely-used treatment methods do effectively change the skills they are designed to target, especially when treatment efficacy research for this population is scarce (Kohnert & Medina, 2009).

Finally, the delayed treatment group did not make significant pre- to post-test change on any English language measure and did not achieve the highest mean group rank for any Aligned-Rank procedure. However, the general trend for the group was
one of improvement on the English language measures. The mean group change from pre- to post-test was in the direction of positive change, or reflected no change, on all English measures. In addition, there are several instances of significant positive individual change, particularly on the EOW-E and the RAN-E. These trends towards improvement in English suggest that school, school-based treatment, and peer interaction likely have a positive effect on the English language skills of at least some bilingual children with PLI. It is also possible that test-retest effects played a role in the observed improvements. The presence of positive trends in this control group created a stringent standard for relative analyses; when the relative change analyses achieved statistical significance, they indicated change above and beyond the positive influences listed here.

Effects of treatment on Spanish language skills.

The final group of research questions related to the individual, absolute, and relative effects of treatment on the Spanish language skills of participants. The only study-related procedures conducted in Spanish were the Spanish language assessments at pre- and post-testing. It was therefore hypothesized that Spanish-language gains attributable to study-related treatment would stem from changes in NCP skills or from cross-linguistic transfer following English gains. Of course, influences on Spanish language exist outside of the study procedures. All children in the study speak mostly Spanish or all Spanish in the home. In addition, the two participants from Jefferson school (#23 and #24 in Table 8 and in Figure 2) were exposed to daily instruction in Spanish language and literacy, as part of Jefferson’s Native Language Literacy
program. Jefferson also employs a bilingual speech-language pathologist who incorporates Spanish into his students’ speech-language treatment sessions.

Across all three groups, none of the absolute change analyses for Spanish reached significance. It is important to note the decreased numbers in the analyses for three of the Spanish measures. Some children did not complete the EOW-S and the ROW-S in the same manner at pre- and at post-testing; these children were excluded from analyses. Several participants could not complete the RAN-S, decreasing the number of participants in those analyses as well.

Despite this consideration, the results suggest that the magnitude of overall Spanish language gains is smaller than overall English language gains. The number of participants in the CELF-4 analyses is the same for both languages, with several significant results in English and none in Spanish. There are several potential explanations for this trend.

First, influences on Spanish growth in this population may be less dominant than the influences on English growth. English is the prestige language in the broader community. It is used in school, in school-based treatment, and in the procedural aspects of study-based treatment. In TD sequential bilingual children who speak a minority home language, an acceleration of growth in the school language is often seen during the school years, alongside a relative plateau in skills in the minority language (e.g., Kohnert & Bates, 2002; Kohnert, Bates, & Hernandez, 1999; Pham, 2011). The steeper trajectory of English skill attainment in the TD population can be attributed to a variety of influences, including the greater social value of English in the broader
community as well as more diverse opportunities to develop this language (such as academic instruction and mass media) (e.g., see Kohnert, 2007, 2010; Pearson, 2007 for discussion). This PLI population has the additional influence of school-based treatment in English administered by an educational speech-language pathologist, and a trend towards relatively greater English growth may simply be consistent with this confluence of influences.

There were two participants at Jefferson whose school experiences also included systematic support for Spanish both in the general classroom and in their school-based treatment. However, these two participants do not demonstrate consistently superior performance on the Spanish language measures administered. Their individual patterns are notably different from each other: participant 23 makes gains on four Spanish measures, while participant 24 has a significant deterioration in performance on six and an improvement in performance on two. In fact, participant 24 demonstrates the greatest drop in performance on Spanish measures of any child in the study (see Figure 2).

This comparison of the two participants from Jefferson leads to a second explanation for the decreased growth in Spanish in comparison to English. Though individual variability in the results has been substantial in all three areas measured here, it appears to be magnified in Spanish. Greater variation in L1 in comparison to L2 is again consistent with studies investigating TD minority language learners in the US (see Kohnert, 2007; 2010 for reviews). In the present study, the appearance of smaller group change in Spanish in comparison to English is driven at least in part by an increase in
the number of individuals who demonstrated a loss of Spanish skills, rather than uniform, but small, growth across the group.

Finally, as noted above, the hypothesized study-related influences on Spanish are all indirect. If the EL treatment is at all effective in changing its targeted skills, it should produce greater change in English than in Spanish. In the NCP group, an effort to minimize language use was made, in order to focus on NCP gains as much as possible. However, the procedural aspects of treatment were conducted in English, and may well have had some impact on English skills. In contrast, Spanish was not used at all in either active treatment condition.

In summary, the lack of significant absolute change in Spanish could result from a combination of less direct study-related influences, less powerful external influences, and individual gains being cancelled out when gains are averaged across the group. Like the absolute change analyses, none of the relative change analyses reached significance. This result indicates that the changes in the active treatment groups did not exceed the changes in the delayed treatment group for any measure. The delayed treatment group did demonstrate the lowest group rank for all Spanish measures except two (ROW-S and CELF-4S Recalling Sentences). Thus, if the current patterns hold in a larger sample, relative treatment effects will appear for Spanish language tasks.

An additional point of interest in the relative change analyses is that some of the relative strengths for each treatment group that were noted in English are maintained in Spanish. The NCP group demonstrates a relative advantage on the Formulated Sentences subtest (with 16.9% of variance accounted for by group membership; \( p = \)
0.14) and the EL group demonstrates a relative advantage on the *Concepts & Following Directions* subtest (with 9.5% of variance accounted for by group membership; \( p = 0.30 \)). These consistencies lend weight to the argument that the two treatments compared here strengthened different skills, independent of language.

However, the EL group’s advantage on *Recalling Sentences* shrinks in Spanish; group membership accounts for 3.1% of variance in scores, \( p = 0.70 \). It is possible that some skills strengthened by the EL treatment (such as auditory memory) transfer easily across languages, whereas others, such as the morphosyntactic knowledge that contributes to *Recalling Sentences* performance, are quite language-specific. This explanation is speculative; these results are small-scale and preliminary for drawing such a conclusion. But cross-linguistic consistencies and inconsistencies will be important clues to the nature of change processes in larger samples.

**Additional Considerations**

In revisiting the research questions, I have considered potential influences on the results in each of the three outcome areas. Several additional factors may affect outcomes across all three categories and deserve further consideration.

**Age.**

The post-hoc analyses found a significant positive correlation between age and change in NCP skills as well as a significant negative correlation between age and change in Spanish language skills. An additional, smaller, positive relationship between age and English language change appeared. The significance of these relationships is not immediately apparent. Participants in the study did span a wide age range; it is
possible that younger children demonstrated greater plasticity in the NCP skills of interest. The literature on treating attention, memory, and processing speed that was reviewed in Chapter 2 does not directly explore the contribution of age to the plasticity of these skills. Similarly, it is possible that younger children were more receptive to English language input, or that the greater NCP change in the younger children fueled greater English language change.

This explanation cannot hold for Spanish, however, in which age was associated with greater change. Another potential explanation stems from the relationship between pre-test score and pre- to post-test change. Pre-test scores were covaried out of the relative change analyses because they can clearly affect the capacity for change. Pre-test scores in English were more closely tied to age than pre-test scores for Spanish, and it is possible that the relationship between age and change in English is mediated by pre-test scores. Replicating the patterns in a larger sample is an important next step towards explaining the relationship.

**Common factors.**

The focus of this study was on the effects of two different treatment programs that were designed to improve specific skills. However, the factors in common across distinct behavioral treatment programs may be more powerful in creating change than the specific ingredients of the treatment programs themselves. This theory, dubbed “Common Factors,” is well-established in counseling psychology (e.g., Wampold, 2001) and emerging in speech-language pathology (Ebert & Kohnert, 2010). This section considers the potential role of some common factors in the treatment results
seen here. Two common factors, treatment satisfaction and treatment intensity, were directly measured in this study. A third, the clinician, was not analyzed, but deserves consideration in future studies.

**Treatment satisfaction.**

This study measured the role of parent and child satisfaction with treatment programs using surveys. The children’s surveys did not differentiate between groups, or even individuals, though they indicated a very high degree of satisfaction with the study. It is possible that a desire to please the person administering the survey also contributed to these results; the survey administrator was not the clinician who administered treatment to the child taking the survey, but the children may nonetheless have perceived the survey administrator to be a person of authority and a representative of the study.

The parent satisfaction surveys also reflected a high degree of overall satisfaction with the project. The group comparison indicated that parents of children in both active treatment groups felt similarly about their experiences. Thus parental attitudes were likely not a factor in any between-group differences found here. However, there was a relationship between total change in English and parental satisfaction with treatment, when the active treatment groups were combined. The correlation does not provide information about the direction of the relationship; it is possible that parental satisfaction was the product of good treatment progress, but it is also possible that good treatment progress was engendered by supportive, enthusiastic
parents. The result does suggest that the effects of treatment go beyond test scores, and that client and family satisfaction is worth exploring in treatment studies.

**Intensity.**

Treatment intensity can be construed as a common factor because it is typically held constant when comparing treatment programs experimentally. It is also a component of treatment structure, a commonly-identified subset of common factors (Grencavage & Norcross, 1990). Though the amount of treatment delivered clearly has the potential to influence the treatment outcomes, there is a dearth of research on this variable in speech-language pathology (Warren, Fey, & Yoder, 2007). In this study, treatment intensity was held as constant as possible: participants in both groups had the same number of treatment sessions, of equal length, scheduled with equal frequency. Treatment was scheduled intensively in an effort to obtain significant changes in a short period of time and to differentiate between school-based treatment and experimental treatment.

The post-hoc analysis of treatment intensity (as defined by the number of sessions attended) was to determine whether within-group differences in session attendance could have influenced the results, rather than to contribute to the broader literature on treatment intensity in speech-language pathology. All children in the two active treatment groups in this study experienced intense treatment. Attendance did not play a significant role in NCP, English, or Spanish change. This statement applies only to children who attended 18 to 24 treatment sessions in this study, as these were the only children analyzed in the active treatment groups. The analysis provides
reassurance that grouping children who attended this range of treatment sessions together for analyses is unlikely to have confounded results. Future studies may be able to gain more insight into the relationship between treatment intensity and treatment gains by analyzing children with a broader range of session attendance scores.

**Clinician effects.**

The clinician or qualities related to the clinician is one of the most frequently-cited common factors. It is highly likely that individual clinicians influence behavioral therapy outcomes (Wampold, 2001). In speech-language pathology, my own past work analyzing clinicians’ self-reported beliefs related to common factors suggests that the influence may be even more individualized: the relationship between a specific client and a specific clinician may impact treatment outcomes (Ebert & Kohnert, 2010). The clinician factor is discussed here because of this importance; individual clinicians, or clinician-client relationships, may well have factored into outcomes.

However, the impact of clinicians was not analyzed here. The ideal method of controlling for clinician influence is to nest participants within clinicians and to consider clinicians to be a random factor when conducting statistical analyses (Wampold, 2001). That solution is not feasible in a study of this modest size. The detailed descriptions of individual clinicians and therapy groups that are provided in Chapter 5 and in Table 6 are intended to acknowledge potential differences and disclose potentially relevant information.
Future Directions

The results of this study provide evidence that NCP treatment changes some NCP skills and also that traditional language treatments change NCP skills. The gains in language skills for both treatments support a link between cognition and language. Furthermore, these treatments may lead to unique patterns of language gains that correspond to both the linguistic and cognitive skills that were strengthened in the respective treatment programs.

Of course, these results are preliminary. The sample here is small and the influences on language growth in bilingual children with PLI are many. Future studies are needed to strengthen, clarify, and expand these findings. Both additional background research and further treatment studies are needed. The purpose of this section is to discuss considerations in planning these future studies.

Time.

An expanded consideration of the role of time in the effects of both NCP and EL treatment would enhance the ability to test this study’s primary hypotheses. In the present study, intensive treatment over a short period of time was used to change skill levels. The time between pre- and post-testing for each participant was a brief 8 weeks. This methodology may be adequate or even advantageous for seeing direct effects of treatment (in this case, seeing effects of EL treatment on English language skills, and effects of NCP treatment on NCP skills).

In contrast, the hypothesis that NCP treatment influences language learning abilities may be better tested using a longer timeframe. If NCP skills underlie language
learning, and improvements in those NCP skills lead to improved language learning, it should require time to begin accumulating additional language knowledge. Changes in language skills such as vocabulary and morphosyntax, as seen on a standardized language test, would likely grow gradually over time. Similarly, cross-linguistic transfer in bilingual children is unlikely to be an instantaneous process.

The practical constraints associated with this study necessitated a short time frame. However, future studies testing this hypothesis would be enhanced by the inclusion of additional testing time points that follow up after the completion of treatment. This follow-up testing would allow the examination of longer-term growth patterns following treatment, allowing a more comprehensive test of the hypothesis.

Task refinement.

Additional refinement of the NCP tasks could enhance the ability to detect skill changes following treatment. Three directions for such research were suggested by the current study. First, longitudinal data for all three tasks is needed. Such data could establish typical patterns seen in repeated testing of school-aged children, both with and without PLI, and thereby clarify the potential contribution of practice effects to post-treatment gains on these NCP tasks.

Secondly, an alternative scoring measure, achievement level, was newly developed in this study for a previously established task, auditory serial memory. The definition of this new variable was supported by the literature and by the data compiled in the current study. It has the additional advantage of accounting for a wide range of performance on this task. However, it has yet to be applied to any other data collected
from the auditory serial memory task in other studies (e.g., Kohnert, Windsor, & Pham, 2009; Yim et al., 2005). An important next step in validating the use of this new measure will be to verify that when applied to these existing databases the achievement level variable accurately reflects performance rather than chance. In addition, this variable should ideally maintain the ability to discriminate between TD and PLI children in linguistically diverse populations, as has been established for the traditional scoring system of the auditory serial memory task (Kohnert, Windsor, & Pham, 2009).

Finally, further refinement of the sustained selective attention task would likely improve the sensitivity of this measure. The task constructed here was based upon two tasks with published data indicating they discriminate between children with and without PLI (Finneran et al., 2009; Spaulding et al., 2008). However, the sample in this study – all of whom had PLI -- demonstrated ceiling effects on the new task. In order for the task to be sensitive across this age range, task difficulty will have to be increased. This may be a complicated goal, as there were two children who could not comprehend and complete the task in its present form. Substantial overall increases in difficulty might serve only to create floor effects for younger or more impaired children. Inserting escalating levels of difficulty, as there are in the auditory serial memory task, may be the best strategy for increasing task sensitivity.

**Individual differences.**

The individual analyses in this study highlighted the individual variability present within all three groups and all three task categories. Though the group trends on each task are almost all positive, almost all tasks have multiple individuals whose
performance deteriorated from pre- to post-testing. Thus, while one obvious way to replicate and extend the current findings would be to conduct a similar group study with larger numbers of participants in each group to increase statistical power, this approach may camouflage individual differences in response to the treatments. A more complete understanding of the relationship between treatment and outcomes could be gained by supplementing group studies with experimental techniques that highlight individual variation. This argument has been raised before in regards to treatment studies in speech-language pathology (e.g., Montgomery & Turkstra, 2003; Pring, 2004); it may be especially powerful for the bilingual PLI population, which demonstrates exceptional variability amongst individuals. Potential methods of accounting for this variability include further use of SSED; large-scale regression analyses to isolate relationships between personal characteristics, treatment features, and outcomes (Wambaugh, 2007); and use of small, homogenous groups in treatment studies to identify precisely which children benefit from a given therapy (Pring, 2004).
Chapter 8: Summary and Conclusions

This study explored the relationship between NCP skills and language skills in bilingual children with PLI by examining the outcomes of treating NCP skills. It was based upon the hypothesis that specific cognitive skills underlie language learning. Processing speed, working memory, and attention are all subtly impaired in the population of children with PLI. Improvements in these three skills might lead to increased language skill in children with PLI, and more specifically to increased language skill in both languages of bilingual children with PLI.

Study participants were assigned to three groups. Two groups completed intensive treatment protocols over approximately 6 weeks and the third completed only pre- and post-testing. Dependent variables spanned three skill areas: NCP, English, and Spanish. Three types of analyses were conducted to examine change from pre- to post-testing for individuals, within groups, and across all three groups.

Study results were mixed with regards to their consistency to the original hypotheses. First, although there was evidence of change for the NCP treatment group in some NCP skills, the NCP group did not demonstrate a relative advantage over other groups in their NCP skill gains immediately following the conclusion of treatment. Instead, the group treated using a traditional English-language protocol demonstrated significant group gains on all three NCP tasks and “outranked” the NCP group in Aligned-Rank procedures for NCP tasks. This result does not indicate that NCP skills are not affected by the NCP treatment, but rather that language-based activities may effectively improve NCP skills like attention and memory.
In English, the group progress for both the NCP and EL treatment groups was consistent with the study hypotheses, which posited that either change in NCP skills or direct treatment of English language targets should be capable of altering English language skills. Furthermore, distinct patterns of change began to appear in the two groups, suggesting that the pathways to English language change differed with the treatments. However, the generally positive overall progress of the delayed treatment control group, coupled with the small number of participants, rendered few of the relative group comparisons statistically significant.

In Spanish, language changes immediately following treatment were not as strong as originally hypothesized. Several factors may have played into this outcome, including outside influences on Spanish growth, individual variation, and an emphasis on English in the setting and procedure of the treatment sessions. There is emerging evidence of Spanish growth in the two active treatment groups, which is consistent with the hypothesis that changes in NCP skills would lead to growth in an untreated language.

This study provides an in-depth exploration of the relationship between cognition and language in children with PLI. Overall, the results support further examination of the changes in speed of processing, attention, and memory following treatment focused on either specific language targets or NCP skills. The evidence here is consistent with a connection between NCP skills and language growth in PLI that may help to create more effective treatments in the future.
### Tables

Table 1.

*Participant characteristics and results from pilot testing of sustained attention task.*

<table>
<thead>
<tr>
<th>Participant</th>
<th>Grade</th>
<th>PLI Status</th>
<th>First Language</th>
<th>ADHD Status</th>
<th>Total Accuracy</th>
<th>Mean RT</th>
<th>SD of RT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>PLI</td>
<td>English</td>
<td>ADHD</td>
<td>178</td>
<td>707.83</td>
<td>130.37</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>PLI</td>
<td>Spanish</td>
<td>TD</td>
<td>195</td>
<td>629.06</td>
<td>127.45</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>PLI</td>
<td>Spanish</td>
<td>TD</td>
<td>195</td>
<td>494.36</td>
<td>129.73</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>PLI</td>
<td>Spanish</td>
<td>TD</td>
<td>196</td>
<td>652.03</td>
<td>100.79</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>PLI</td>
<td>Spanish</td>
<td>TD</td>
<td>197</td>
<td>516.24</td>
<td>112.48</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>PLI</td>
<td>Spanish</td>
<td>TD</td>
<td>197</td>
<td>459.82</td>
<td>125.81</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>TD</td>
<td>Spanish</td>
<td>TD</td>
<td>200</td>
<td>631.77</td>
<td>93.01</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>PLI</td>
<td>Spanish</td>
<td>TD</td>
<td>197</td>
<td>462.11</td>
<td>119.97</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>PLI</td>
<td>English</td>
<td>ADHD</td>
<td>174</td>
<td>536.57</td>
<td>157.08</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>PLI</td>
<td>Spanish</td>
<td>TD</td>
<td>143</td>
<td>648.81</td>
<td>142.23</td>
</tr>
<tr>
<td>11</td>
<td>4</td>
<td>PLI</td>
<td>English</td>
<td>ADHD</td>
<td>196</td>
<td>670.19</td>
<td>103.84</td>
</tr>
<tr>
<td>12</td>
<td>5</td>
<td>TD</td>
<td>Spanish</td>
<td>TD</td>
<td>199</td>
<td>476.77</td>
<td>110.55</td>
</tr>
<tr>
<td>13</td>
<td>5</td>
<td>TD</td>
<td>Hmong</td>
<td>ADHD</td>
<td>195</td>
<td>492.34</td>
<td>121.51</td>
</tr>
<tr>
<td>14</td>
<td>4</td>
<td>TD</td>
<td>English</td>
<td>ADHD</td>
<td>195</td>
<td>447</td>
<td>146.83</td>
</tr>
</tbody>
</table>

*Note.* Total accuracy = number of correct responses out of 200; mean RT = mean reaction time for correct responses, after trimming outliers more than 2 standard deviations from the individual mean RT; SD of RT = standard deviation of mean reaction time for correct responses, after the trimming procedure.
Table 2.

*Repeated dependent measures and results from preliminary SSED studies* (Ebert & Kohnert, 2009; Ebert, Rentmeester, & Kohnert, 2011).

<table>
<thead>
<tr>
<th>Task</th>
<th>Description</th>
<th>Effect size (d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Choice Visual Detection</td>
<td>Respond as quickly as possible when a red or blue circle appears on screen</td>
<td>-0.19 0.74 1.67 1.81</td>
</tr>
<tr>
<td>Auditory Serial Memory</td>
<td>Remember pairs of tone patterns and judge same/different</td>
<td>0.12 -0.14 NA NA</td>
</tr>
<tr>
<td>Rapid Automatic Naming (Eng)</td>
<td>Name color/shape combinations as quickly as possible (English)</td>
<td>1.86 1.85 2.30 0.88</td>
</tr>
<tr>
<td>Nonword Repetition (Eng)</td>
<td>Recall and reproduce nonsense words that follow English phonotactic constraints</td>
<td>-0.55 1.28 2.06 1.27</td>
</tr>
<tr>
<td>Nonword Repetition (Span)</td>
<td>Recall and reproduce nonsense words that follow Spanish phonotactic constraints</td>
<td>NA NA 2.01 1.12</td>
</tr>
<tr>
<td>Sentence Repetition (Eng)</td>
<td>Repeat sentences of varying length and complexity in English</td>
<td>NA NA 2.03 1.80</td>
</tr>
</tbody>
</table>

*Note.* Effect sizes are calculated by subtracting the mean of the final three treatment data points from the mean of the baseline data points, and dividing the result by pooled standard deviation for that measure (Gillam et al., 2001). Effect sizes greater than 0.8 are considered clinically significant change (Gillam et al., 2001). Negative effect sizes indicate poorer performance at the end of treatment. P1 and P2 were monolingual participants in the first preliminary SSED study (Ebert & Kohnert, 2009); P3 and P4 were bilingual participants in the second preliminary SSED study (Ebert, Rentmeester, & Kohnert, 2011).
Table 3.

*Nonlinguistic cognitive processing measures administered at pre- and post-testing.*

<table>
<thead>
<tr>
<th>Name</th>
<th>Task</th>
<th>Primary Dependent Variable</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auditory Serial Memory</td>
<td>Remember pairs of tone patterns and judge same/different</td>
<td>Highest level of difficulty achieved</td>
<td>Yim, 2006</td>
</tr>
<tr>
<td>Choice Visual Detection</td>
<td>Respond as quickly as possible when a red or blue circle appears on screen</td>
<td>RT for accurate trials</td>
<td>Kohnert &amp; Windsor, 2004</td>
</tr>
<tr>
<td>Sustained Selective Attention</td>
<td>Attend to a stream of sounds and respond only to target sounds</td>
<td>d'</td>
<td>Developed from Finneran et al., 2009; Spaulding et al., 2008</td>
</tr>
</tbody>
</table>
Table 4.

*English and Spanish language measures administered at pre- and post-testing.*

<table>
<thead>
<tr>
<th>English name and source</th>
<th>Spanish name and source</th>
<th>Construct or skill measured</th>
<th>Dependent variable(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CELF-4E (Semel et al., 2003)</td>
<td>CELF-4S (Wiig et al., 2006)</td>
<td>Global language, including expressive and receptive semantic and syntactic skills</td>
<td>Core language standardized score + component subtest scaled scores + raw scores</td>
</tr>
<tr>
<td>EOW-E (Brownell, 2000a)</td>
<td>EOW-S (Brownell, 2001a)</td>
<td>Expressive vocabulary</td>
<td>Standardized score + raw score</td>
</tr>
<tr>
<td>ROW-E (Brownell, 2000b)</td>
<td>ROW-S (Brownell, 2001b)</td>
<td>Receptive vocabulary</td>
<td>Standardized score + raw score</td>
</tr>
<tr>
<td>NWR-E (Dollaghan &amp; Campbell, 1998)</td>
<td>NWR-S (Ebert et al., 2008)</td>
<td>Recall and production of nonsense words following language-specific constraints</td>
<td>Percent phonemes correct across all word lengths</td>
</tr>
<tr>
<td>RAN from CELF-4E (Semel et al., 2003)</td>
<td>RAN from CELF-4S (Wiig et al., 2006)</td>
<td>Rapidly naming color-shape combinations in sequence</td>
<td>Ratio of time in seconds to number of accurate responses</td>
</tr>
</tbody>
</table>

*Note.* Abbreviations are as follows: CELF-4E = Clinical Evaluation of Language Fundamentals-4th Edition, English, Core Language Score; CELF-4S = Clinical Evaluation of Language Fundamentals-4th Edition, Spanish, Core Language Score; EOW-E = Expressive One-Word Vocabulary Test (English); EOW-S = Expressive One-Word Vocabulary Test, Bilingual Edition (Spanish); ROW-E = Receptive One-Word Vocabulary Test (English); ROW-S = Receptive One-Word Vocabulary Test, Bilingual Edition (Spanish); NWR-E = English Nonword Repetition; NWR-S = Spanish Nonword Repetition; RAN = Rapid Automatic Naming.
Table 5.

*Participant characteristics and pretesting scores by treatment condition.*

<table>
<thead>
<tr>
<th>Variable</th>
<th>NCP ( (n = 10) )</th>
<th>EL ( (n = 8) )</th>
<th>DEL ( (n = 6) )</th>
<th>Results of Group Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>100%</td>
<td>50%</td>
<td>83.3%</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>98.3 (11.6)</td>
<td>94.1 (15.0)</td>
<td>93.0 (11.6)</td>
<td>( H = 1.43, p = 0.49 )</td>
</tr>
<tr>
<td>TONI</td>
<td>100.4 (19.1)</td>
<td>92.0 (10.3)</td>
<td>88.3 (8.6)</td>
<td>( H = 3.79, p = 0.15 )</td>
</tr>
<tr>
<td>CELF-4E</td>
<td>49.4 (10.8)</td>
<td>51.1 (12.0)</td>
<td>42.3 (4.1)</td>
<td>( H = 2.65, p = 0.27 )</td>
</tr>
<tr>
<td>CELF-4S</td>
<td>59.0 (11.9)</td>
<td>64.6 (13.5)</td>
<td>53.0 (10.3)</td>
<td>( H = 2.80, p = 0.25 )</td>
</tr>
<tr>
<td>EOW-E</td>
<td>64.6 (6.9)</td>
<td>64.9 (9.0)</td>
<td>61.7 (5.6)</td>
<td>( H = 0.84, p = 0.66 )</td>
</tr>
<tr>
<td>EOW-S</td>
<td>63.0 (9.4) ( [n = 5] )</td>
<td>64.0 (5.7) ( [n = 2] )</td>
<td>73.5 (26.2) ( [n = 2] )</td>
<td>( H = 0.25, p = 0.88 )</td>
</tr>
<tr>
<td>ROW-E</td>
<td>76.9 (4.4)</td>
<td>76.0 (4.8)</td>
<td>73.0 (7.5)</td>
<td>( H = 0.91, p = 0.64 )</td>
</tr>
<tr>
<td>ROW-S</td>
<td>75.8 (10.6) ( [n = 5] )</td>
<td>71.0 (9.9) ( [n = 2] )</td>
<td>95.0 (NA) ( [n = 1] )</td>
<td>( H = 2.63, p = 0.27 )</td>
</tr>
<tr>
<td>NWR-E</td>
<td>65.5 (15.1) ( [n = 9] )</td>
<td>66.3 (13.2)</td>
<td>61.1 (17.2)</td>
<td>( H = 0.51, p = 0.78 )</td>
</tr>
<tr>
<td>NWR-S</td>
<td>71.5 (17.5)</td>
<td>72.3 (22.7)</td>
<td>69.1 (26.7)</td>
<td>( H = 0.12, p = 0.94 )</td>
</tr>
<tr>
<td>RAN-E</td>
<td>1.84 (1.2)</td>
<td>2.07 (1.0)</td>
<td>2.50 (1.7)</td>
<td>( H = 1.19, p = 0.55 )</td>
</tr>
<tr>
<td>RAN-S</td>
<td>1.71 (0.5) ( [n = 7] )</td>
<td>2.66 (1.5) ( [n = 7] )</td>
<td>3.49 (2.4)</td>
<td>( H = 3.27, p = 0.20 )</td>
</tr>
<tr>
<td>CVD</td>
<td>719.4 (85.5)</td>
<td>807.5 (145.8)</td>
<td>691.3 (103.2)</td>
<td>( H = 3.48, p = 0.18 )</td>
</tr>
<tr>
<td>ASM</td>
<td>2.30 (1.7)</td>
<td>2.13 (1.9)</td>
<td>1.33 (1.5)</td>
<td>( H = 1.59, p = 0.45 )</td>
</tr>
<tr>
<td>SSA</td>
<td>3.54 (1.1)</td>
<td>2.93 (1.3)</td>
<td>3.35 (1.6)</td>
<td>( H = 1.20, p = 0.55 )</td>
</tr>
</tbody>
</table>
Note. Group comparisons were conducted using the Kruskal-Wallis test. The resulting $H$ statistic follows a Chi-squared distribution, with 2 degrees of freedom. Exceptions to the group $n$s are reported inside brackets in individual cells. Descriptive statistics and the group comparisons for Spanish vocabulary tests (EOW-S and ROW-S) were calculated using only scores from testing conducted entirely in Spanish. Abbreviations are as follows: NCP = nonlinguistic cognitive processing treatment group; EL = English language treatment group; DEL = delayed treatment group; TONI = Test of Nonverbal Intelligence-3rd Edition; CVD = Choice Visual Detection; ASM = Auditory Serial Memory; SSA = Sustained Selective Attention. See Table 4 note for additional abbreviations. All test scores are standard scores and are reported as Mean (SD). Gender is reported as percentage of males in the group. Age is reported in months and indicates the participant’s age at the time of pre-testing. NWR scores are reported as total Percent Phonemes Correct. RAN scores are reported as the ratio of the time in seconds to the number of correct responses. CVD = mean RT from Choice Visual Detection task; ASM = highest level of achievement on Auditory Serial Memory task; SSA = $d'$ score from Sustained Selective Attention task, corrected for late hits.
Table 6.

*Description of treatment groups.*

<table>
<thead>
<tr>
<th>Group</th>
<th>Size</th>
<th>Treatment type</th>
<th>Cycle</th>
<th>Clinician</th>
<th>School Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>NCP</td>
<td>1</td>
<td>A</td>
<td>Green Central</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>EL</td>
<td>1</td>
<td>B</td>
<td>Green Central</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>EL</td>
<td>2</td>
<td>C</td>
<td>Whittier</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>NCP</td>
<td>2</td>
<td>B</td>
<td>Green Central</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>NCP</td>
<td>2</td>
<td>A</td>
<td>Armatage Montessori</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>NCP</td>
<td>3</td>
<td>A</td>
<td>Andersen United</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>NCP</td>
<td>3</td>
<td>B</td>
<td>Andersen United</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>EL</td>
<td>3</td>
<td>D</td>
<td>Jefferson</td>
</tr>
</tbody>
</table>

*Note.* Group size = number of participants who attended the group; NCP = nonlinguistic cognitive processing treatment group; EL = English language treatment group; DEL = delayed treatment group.
Table 7.

*Summary of treatment fidelity review.*

<table>
<thead>
<tr>
<th>Group</th>
<th>Sessions coded</th>
<th>Treatment time</th>
<th>Spanish</th>
<th>Irrelevant</th>
<th>Redirect</th>
<th>Linguistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCP</td>
<td>21 (18.4%)</td>
<td>76.2 (10.8)</td>
<td>0 (0)</td>
<td>0.7 (0.2)</td>
<td>8.6 (7.2)</td>
<td>0.2 (0.5)</td>
</tr>
<tr>
<td>EL</td>
<td>13 (18.8%)</td>
<td>67.7 (5.2)</td>
<td>0.1 (0.2)</td>
<td>1.3 (2.0)</td>
<td>11.2 (10.8)</td>
<td>16.4 (7.3)</td>
</tr>
</tbody>
</table>

*Note.* Sessions coded reports the total number of treatment sessions coded for treatment fidelity purposes, with the percentage of total sessions coded in parentheses. Treatment time indicates the total number of minutes a single child spent completing treatment tasks in a session, and is reported as mean (SD). Spanish, irrelevant, redirect, and linguistic report the number of comments in the given category that were directed to a single participant in a single session. These categories are also reported as mean (SD).
Table 8a.

**Individual change on English language measures.**

<table>
<thead>
<tr>
<th>ID</th>
<th>CY</th>
<th>GRP</th>
<th>CD</th>
<th>WS</th>
<th>WC</th>
<th>RS</th>
<th>FS</th>
<th>CLF</th>
<th>RAN</th>
<th>NWR</th>
<th>EOW</th>
<th>ROW</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>NCP</td>
<td>0</td>
<td>NA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>NCP</td>
<td>+</td>
<td>0</td>
<td>NA</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>NCP</td>
<td>+</td>
<td>0</td>
<td>NA</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>–</td>
<td>0</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>NCP</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>X</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>NCP</td>
<td>+</td>
<td>0</td>
<td>NA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>−</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>3</td>
<td>NCP</td>
<td>0</td>
<td>NA</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>−</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>3</td>
<td>NCP</td>
<td>−</td>
<td>0</td>
<td>NA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>−</td>
<td>0</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>3</td>
<td>NCP</td>
<td>−</td>
<td>NA</td>
<td>+</td>
<td>−</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>21</td>
<td>3</td>
<td>NCP</td>
<td>+</td>
<td>+</td>
<td>NA</td>
<td>+</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>−</td>
</tr>
<tr>
<td>22</td>
<td>3</td>
<td>NCP</td>
<td>−</td>
<td>NA</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>EL</td>
<td>0</td>
<td>NA</td>
<td>0</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>EL</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>EL</td>
<td>+</td>
<td>0</td>
<td>NA</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>2</td>
<td>EL</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>EL</td>
<td>0 +</td>
<td>NA</td>
<td>+</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>EL</td>
<td>0</td>
<td>NA</td>
<td>−</td>
<td>+</td>
<td>0</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>3</td>
<td>EL</td>
<td>+</td>
<td>+</td>
<td>NA</td>
<td>+</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>3</td>
<td>EL</td>
<td>+</td>
<td>0</td>
<td>NA</td>
<td>+</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>−</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>DEL</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>X</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>DEL</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>−</td>
<td>X</td>
<td>0</td>
<td>−</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>DEL</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>−</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>DEL</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>17</td>
<td>2</td>
<td>DEL</td>
<td>+</td>
<td>X</td>
<td>NA</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>3</td>
<td>DEL</td>
<td>−</td>
<td>X</td>
<td>NA</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>+</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Each row in the table represents data from a single participant across all English language measures. + indicates the participant’s pre- to post-test change score was positive and met or exceeded the task criterion; − indicates the participant’s pre- to post-test change score was negative and met or exceeded the task criterion; and 0 indicates the participant’s change score met neither of the above conditions. X indicates missing data; specifically, one NWR sample for participant 7 was lost, and NWR testing for participant 3 in cycle 1 is invalid as control data due to 3’s participation in phonological treatment during that time period. In addition, the WS subtest CELF-4 subtest was inconsistent from pre- to post-testing for 17 and 16 because these children turned 9 between pre- and post-testing. Abbreviations: ID = anonymous identification code for
an individual participant; CY = treatment cycle; GRP = group; NCP = nonlinguistic cognitive processing treatment; EL = English language treatment; DEL = delayed treatment; CD = *Concepts and Following Directions* subtest of CELF-4E; WS = *Word Structure* subtest of CELF-4E for children aged 6;0-8;11; WC = *Word Classes* subtest of CELF-4E for children aged 9;0-9;11; RS = *Recalling Sentences* subtest of CELF-4E; FS = *Formulated Sentences* subtest of CELF-4E; CLF = Core Language score from CELF-4E.
### Table 8b.

**Individual change on Spanish language measures.**

<table>
<thead>
<tr>
<th>ID</th>
<th>CY</th>
<th>GRP</th>
<th>CD</th>
<th>WS</th>
<th>WC</th>
<th>RS</th>
<th>FS</th>
<th>CLF</th>
<th>RAN</th>
<th>NWR</th>
<th>EOW</th>
<th>ROW</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>NCP</td>
<td>0 NA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>X^1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>NCP</td>
<td>+</td>
<td>0</td>
<td>NA</td>
<td>−</td>
<td>0</td>
<td>0</td>
<td>X^2</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>NCP</td>
<td>−</td>
<td>0</td>
<td>NA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>X^2</td>
<td>0</td>
<td>−</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>NCP</td>
<td>0</td>
<td>+</td>
<td>NA</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>−</td>
<td></td>
<td></td>
<td>X^1</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>NCP</td>
<td>0</td>
<td>+</td>
<td>NA</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td></td>
<td>X^1</td>
</tr>
<tr>
<td>17</td>
<td>3</td>
<td>NCP</td>
<td>0</td>
<td>NA</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>3</td>
<td>NCP</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>−</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>3</td>
<td>NCP</td>
<td>+</td>
<td>NA</td>
<td>+</td>
<td>0</td>
<td>−</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td></td>
<td>−</td>
</tr>
<tr>
<td>21</td>
<td>3</td>
<td>NCP</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>22</td>
<td>3</td>
<td>NCP</td>
<td>−</td>
<td>NA</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>+</td>
<td></td>
<td>−</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>EL</td>
<td>+</td>
<td>NA</td>
<td>+</td>
<td>0</td>
<td>−</td>
<td>+</td>
<td>X^2</td>
<td>−</td>
<td></td>
<td>X^1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>EL</td>
<td>+</td>
<td>0</td>
<td>NA</td>
<td>−</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>+</td>
<td></td>
<td>−</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>EL</td>
<td>−</td>
<td>+</td>
<td>NA</td>
<td>0</td>
<td>0</td>
<td>−</td>
<td>0</td>
<td>+</td>
<td></td>
<td>−</td>
</tr>
<tr>
<td>13</td>
<td>2</td>
<td>EL</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>−</td>
<td>0</td>
<td></td>
<td>X^1</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>EL</td>
<td>+</td>
<td>0</td>
<td>NA</td>
<td>+</td>
<td>−</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>EL</td>
<td>+</td>
<td>NA</td>
<td>0</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>3</td>
<td>EL</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>+</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>24</td>
<td>3</td>
<td>EL</td>
<td>−</td>
<td>+</td>
<td>NA</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>X^2</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>DEL</td>
<td>+</td>
<td>−</td>
<td>NA</td>
<td>−</td>
<td>0</td>
<td>0</td>
<td>X^2</td>
<td>X^3</td>
<td>X^1</td>
<td>X^1</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>DEL</td>
<td>+</td>
<td>0</td>
<td>NA</td>
<td>0</td>
<td>−</td>
<td>0</td>
<td>X^2</td>
<td>+</td>
<td>0</td>
<td>−</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>DEL</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>−</td>
<td>+</td>
<td>0</td>
<td>+</td>
<td>−</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>DEL</td>
<td>+</td>
<td>0</td>
<td>NA</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>17</td>
<td>2</td>
<td>DEL</td>
<td>+</td>
<td>X^1</td>
<td>NA</td>
<td>−</td>
<td>0</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td></td>
<td>X^1</td>
</tr>
<tr>
<td>16</td>
<td>3</td>
<td>DEL</td>
<td>−</td>
<td>X^1</td>
<td>NA</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td></td>
<td>−</td>
</tr>
</tbody>
</table>

**Note.** Each row in the table represents data from a single participant across all Spanish language measures. Individual change ratings for Spanish ROW and EOW are included only if the test was conducted in the consistent manner at pre- and post-testing for that participant; if not, an X for missing data is indicated. Specific missing data codes include: X^1 = data missing due to inconsistent administration from pre- to post-testing; X^2 = data missing because participant was unable to complete task at either pre- or post-testing; X^3 = data not included because participant completed phonological treatment during cycle, impacting NWR. See Table 8a note for abbreviations and additional details.
Table 8c.

**Individual change on nonlinguistic cognitive processing measures.**

<table>
<thead>
<tr>
<th>ID</th>
<th>CY</th>
<th>GRP</th>
<th>CVD</th>
<th>ASM</th>
<th>SSA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>NCP</td>
<td>+</td>
<td>0</td>
<td>−</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>NCP</td>
<td>+</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>NCP</td>
<td>0</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>NCP</td>
<td>0</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>NCP</td>
<td>0</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>17</td>
<td>3</td>
<td>NCP</td>
<td>0</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>18</td>
<td>3</td>
<td>NCP</td>
<td>0</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>19</td>
<td>3</td>
<td>NCP</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>21</td>
<td>3</td>
<td>NCP</td>
<td>+</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>22</td>
<td>3</td>
<td>NCP</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>EL</td>
<td>0</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>EL</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>EL</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>13</td>
<td>2</td>
<td>EL</td>
<td>0</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>EL</td>
<td>0</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>EL</td>
<td>+</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>23</td>
<td>3</td>
<td>EL</td>
<td>0</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>24</td>
<td>3</td>
<td>EL</td>
<td>0</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>DEL</td>
<td>0</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>DEL</td>
<td>+</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>DEL</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>DEL</td>
<td>0</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>17</td>
<td>2</td>
<td>DEL</td>
<td>0</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>16</td>
<td>3</td>
<td>DEL</td>
<td>0</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

*Note.* Each row in the table represents data from a single participant across all nonlinguistic cognitive processing measures. See note for Table 8a. Additional abbreviations: CVD = choice visual detection; ASM = auditory serial memory; SSA = sustained selective attention.
Table 9a.

**Results of Fisher matched-pairs exact test comparing pre-test to post-test scores for nonlinguistic cognitive processing treatment group.**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Pre M</th>
<th>SD</th>
<th>Post M</th>
<th>SD</th>
<th>Mean Change</th>
<th>d</th>
<th>Perm</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASM</td>
<td>2.30</td>
<td>1.70</td>
<td>2.60</td>
<td>1.50</td>
<td>0.30</td>
<td>0.20</td>
<td>384/1024</td>
<td>0.38</td>
</tr>
<tr>
<td>CVD</td>
<td>719.4</td>
<td>85.5</td>
<td>652.8</td>
<td>77.9</td>
<td>-66.8</td>
<td>0.86</td>
<td>36/1024</td>
<td>0.04*</td>
</tr>
<tr>
<td>SSA</td>
<td>3.54</td>
<td>1.12</td>
<td>3.80</td>
<td>0.78</td>
<td>0.26</td>
<td>0.28</td>
<td>105/1024</td>
<td>0.10</td>
</tr>
<tr>
<td>ROW-E</td>
<td>76.9</td>
<td>4.4</td>
<td>79.6</td>
<td>3.7</td>
<td>2.7</td>
<td>0.70</td>
<td>89/1024</td>
<td>0.09</td>
</tr>
<tr>
<td>EOW-E</td>
<td>64.6</td>
<td>6.9</td>
<td>66.7</td>
<td>9.5</td>
<td>2.1</td>
<td>0.27</td>
<td>128/1024</td>
<td>0.13</td>
</tr>
<tr>
<td>CELF-CD-E</td>
<td>2.5</td>
<td>1.9</td>
<td>2.5</td>
<td>1.4</td>
<td>0.0</td>
<td>0</td>
<td>588/1024</td>
<td>0.57</td>
</tr>
<tr>
<td>CELF-RS-E</td>
<td>1.6</td>
<td>1.3</td>
<td>1.8</td>
<td>1.3</td>
<td>0.2</td>
<td>0.16</td>
<td>320/1024</td>
<td>0.31</td>
</tr>
<tr>
<td>CELF-FS-E</td>
<td>2.1</td>
<td>1.6</td>
<td>3.3</td>
<td>2.9</td>
<td>1.2</td>
<td>0.54</td>
<td>32/1024</td>
<td>0.03*</td>
</tr>
<tr>
<td>CELF-CO-E</td>
<td>49.4</td>
<td>10.8</td>
<td>53.7</td>
<td>11.2</td>
<td>4.3</td>
<td>0.41</td>
<td>8/1024</td>
<td>0.008**</td>
</tr>
<tr>
<td>RAN-E</td>
<td>1.84</td>
<td>1.22</td>
<td>1.67</td>
<td>0.96</td>
<td>-0.17</td>
<td>0.16</td>
<td>202/1024</td>
<td>0.20</td>
</tr>
<tr>
<td>NWR-E</td>
<td>65.5</td>
<td>15.1</td>
<td>63.2</td>
<td>10.2</td>
<td>-2.3</td>
<td>-0.19</td>
<td>376/512</td>
<td>0.73</td>
</tr>
<tr>
<td>ROW-S</td>
<td>71.3</td>
<td>11.9</td>
<td>79.7</td>
<td>16.6</td>
<td>8.4</td>
<td>0.63</td>
<td>23/128</td>
<td>0.18</td>
</tr>
<tr>
<td>EOW-S</td>
<td>67.1</td>
<td>11.0</td>
<td>71.8</td>
<td>12.7</td>
<td>4.7</td>
<td>0.42</td>
<td>48/512</td>
<td>0.09</td>
</tr>
<tr>
<td>CELF-CD-S</td>
<td>4.2</td>
<td>1.7</td>
<td>4.82</td>
<td>1.9</td>
<td>0.0</td>
<td>0</td>
<td>560/1024</td>
<td>0.55</td>
</tr>
<tr>
<td>CELF-RS-S</td>
<td>3.3</td>
<td>1.8</td>
<td>3.6</td>
<td>1.5</td>
<td>0.3</td>
<td>0.19</td>
<td>272/1024</td>
<td>0.27</td>
</tr>
<tr>
<td>CELF-FS-S</td>
<td>3.7</td>
<td>2.3</td>
<td>4.0</td>
<td>3.1</td>
<td>0.3</td>
<td>0.12</td>
<td>336/1024</td>
<td>0.33</td>
</tr>
<tr>
<td>CELF-CO-S</td>
<td>59.0</td>
<td>11.9</td>
<td>60.9</td>
<td>11.3</td>
<td>1.9</td>
<td>0.17</td>
<td>68/1024</td>
<td>0.07</td>
</tr>
<tr>
<td>RAN-S</td>
<td>1.71</td>
<td>0.46</td>
<td>1.58</td>
<td>0.43</td>
<td>-0.13</td>
<td>0.32</td>
<td>29/128</td>
<td>0.23</td>
</tr>
<tr>
<td>NWR-S</td>
<td>71.5</td>
<td>17.5</td>
<td>77.8</td>
<td>13.8</td>
<td>6.3</td>
<td>0.42</td>
<td>69/1024</td>
<td>0.07</td>
</tr>
</tbody>
</table>

*Note.* The mean and standard deviation for the group pre- and post-test scores on the primary outcome measure for each task are reported. See note for Table 5 for a list of
the primary outcome measure for each task; standard scores were used for all standardized language assessments. The mean change column reports the group mean change score; positive change scores represent growth for all tasks except the speed tasks: CVD, RAN-E, and RAN-S. Effect sizes are reported under $d$; negative signs on the effect sizes indicate that group task performance became worse from pre- to post-testing. The Perm (permutations) column reports the number of potential mean change scores that would be equal to or more extreme than the observed change score, divided by all possible mean change scores. The $p$ column reports one-tailed probability of the observed change score occurring by chance. *$p < 0.05$, **$p < 0.01$. 
Table 9b.

Results of Fisher matched-pairs exact test comparing pre-test to post-test scores for English language treatment group.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Pre M, SD</th>
<th>Post M, SD</th>
<th>Mean Change</th>
<th>d</th>
<th>Perm</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASM</td>
<td>2.13, 1.89</td>
<td>3.13, 1.46</td>
<td>1.0</td>
<td>0.63</td>
<td>12/256</td>
<td>0.047*</td>
</tr>
<tr>
<td>CVD</td>
<td>807.5, 145.8</td>
<td>691.5, 94.9</td>
<td>-116.0</td>
<td>1.01</td>
<td>6/256</td>
<td>0.02*</td>
</tr>
<tr>
<td>SSA</td>
<td>2.93, 1.30</td>
<td>3.81, 0.73</td>
<td>0.87</td>
<td>0.90</td>
<td>1/128</td>
<td>0.008**</td>
</tr>
<tr>
<td>ROW-E</td>
<td>76.0, 4.8</td>
<td>76.5, 10.1</td>
<td>0.5</td>
<td>0.07</td>
<td>115/256</td>
<td>0.45</td>
</tr>
<tr>
<td>EOW-E</td>
<td>64.9, 9.0</td>
<td>72.9, 11.0</td>
<td>8.0</td>
<td>0.85</td>
<td>2/256</td>
<td>0.008**</td>
</tr>
<tr>
<td>CELF-CD-E</td>
<td>2.8, 2.0</td>
<td>3.9, 2.0</td>
<td>1.1</td>
<td>0.59</td>
<td>32/256</td>
<td>0.13</td>
</tr>
<tr>
<td>CELF-RS-E</td>
<td>1.9, 1.3</td>
<td>3.0, 1.9</td>
<td>1.1</td>
<td>0.72</td>
<td>4/256</td>
<td>0.02*</td>
</tr>
<tr>
<td>CELF-FS-E</td>
<td>2.4, 1.8</td>
<td>2.9, 2.6</td>
<td>0.5</td>
<td>0.24</td>
<td>64/256</td>
<td>0.25</td>
</tr>
<tr>
<td>CELF-CO-E</td>
<td>51.1, 12.0</td>
<td>56.6, 13.9</td>
<td>5.5</td>
<td>0.45</td>
<td>8/256</td>
<td>0.03*</td>
</tr>
<tr>
<td>RAN-E</td>
<td>2.07, 1.0</td>
<td>1.57, 0.43</td>
<td>-0.5</td>
<td>0.69</td>
<td>15/256</td>
<td>0.06</td>
</tr>
<tr>
<td>NWR-E</td>
<td>66.3, 13.2</td>
<td>72.1, 10.9</td>
<td>5.8</td>
<td>0.51</td>
<td>36/256</td>
<td>0.14</td>
</tr>
<tr>
<td>ROW-S</td>
<td>87.8, 17.1</td>
<td>82.6, 17.1</td>
<td>-5.2</td>
<td>0.34</td>
<td>26/32</td>
<td>0.81</td>
</tr>
<tr>
<td>EOW-S</td>
<td>81.0, 13.1</td>
<td>82.5, 20.3</td>
<td>1.5</td>
<td>0.09</td>
<td>97/256</td>
<td>0.38</td>
</tr>
<tr>
<td>CELF-CD-S</td>
<td>5.4, 3.3</td>
<td>6.6, 2.1</td>
<td>1.3</td>
<td>0.46</td>
<td>44/256</td>
<td>0.17</td>
</tr>
<tr>
<td>CELF-RS-S</td>
<td>4.5, 2.1</td>
<td>4.5, 2.8</td>
<td>0.0</td>
<td>0</td>
<td>156/256</td>
<td>0.61</td>
</tr>
<tr>
<td>CELF-FS-S</td>
<td>4.9, 2.1</td>
<td>4.5, 2.1</td>
<td>-0.4</td>
<td>0.20</td>
<td>164/256</td>
<td>0.64</td>
</tr>
<tr>
<td>CELF-CO-S</td>
<td>64.6, 13.5</td>
<td>68.9, 14.8</td>
<td>4.3</td>
<td>0.32</td>
<td>17/256</td>
<td>0.07</td>
</tr>
<tr>
<td>RAN-S</td>
<td>2.34, 1.33</td>
<td>1.93, 0.82</td>
<td>-0.41</td>
<td>0.41</td>
<td>25/64</td>
<td>0.39</td>
</tr>
<tr>
<td>NWR-S</td>
<td>72.4, 22.7</td>
<td>83.9, 6.1</td>
<td>11.5</td>
<td>0.74</td>
<td>32/256</td>
<td>0.13</td>
</tr>
</tbody>
</table>

*Note.* See Table 9a note.
Table 9c.

*Results of Fisher matched-pairs exact test comparing pre-test to post-test scores for delayed treatment group.*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Pre</th>
<th>Post</th>
<th>Mean</th>
<th>Change</th>
<th>d</th>
<th>Perm</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASM</td>
<td>1.33</td>
<td>1.51</td>
<td>2.00</td>
<td>2.80</td>
<td>0.7</td>
<td>0.33</td>
<td>16/64</td>
</tr>
<tr>
<td>CVD</td>
<td>691.3</td>
<td>103.2</td>
<td>650</td>
<td>33.0</td>
<td>-40.9</td>
<td>0.60</td>
<td>8/32</td>
</tr>
<tr>
<td>SSA</td>
<td>3.36</td>
<td>1.56</td>
<td>3.90</td>
<td>1.56</td>
<td>0.55</td>
<td>0.39</td>
<td>2/32</td>
</tr>
<tr>
<td>ROW-E</td>
<td>73.0</td>
<td>7.5</td>
<td>74.5</td>
<td>5.2</td>
<td>1.5</td>
<td>0.25</td>
<td>23/64</td>
</tr>
<tr>
<td>EOW-E</td>
<td>61.7</td>
<td>5.6</td>
<td>66.0</td>
<td>9.4</td>
<td>4.3</td>
<td>0.61</td>
<td>8/64</td>
</tr>
<tr>
<td>CELF-CD-E</td>
<td>1.5</td>
<td>1.2</td>
<td>1.5</td>
<td>0.8</td>
<td>0.0</td>
<td>0</td>
<td>48/64</td>
</tr>
<tr>
<td>CELF-RS-E</td>
<td>1.0</td>
<td>0.0</td>
<td>1.2</td>
<td>0.4</td>
<td>0.2</td>
<td>0.77</td>
<td>32/64</td>
</tr>
<tr>
<td>CELF-FS-E</td>
<td>1.2</td>
<td>0.2</td>
<td>1.2</td>
<td>0.2</td>
<td>0.0</td>
<td>0.69</td>
<td>64/64</td>
</tr>
<tr>
<td>CELF-CO-E</td>
<td>42.3</td>
<td>4.1</td>
<td>44.0</td>
<td>6.3</td>
<td>1.7</td>
<td>0.17</td>
<td>16/64</td>
</tr>
<tr>
<td>RAN-E</td>
<td>2.50</td>
<td>1.67</td>
<td>2.13</td>
<td>1.63</td>
<td>-0.38</td>
<td>0.25</td>
<td>10/64</td>
</tr>
<tr>
<td>NWR-E</td>
<td>66.5</td>
<td>12.6</td>
<td>67.5</td>
<td>10.5</td>
<td>1.0</td>
<td>0.10</td>
<td>16/32</td>
</tr>
<tr>
<td>ROW-S</td>
<td>97.5</td>
<td>8.7</td>
<td>94.8</td>
<td>13.2</td>
<td>-2.8</td>
<td>-0.28</td>
<td>10/16</td>
</tr>
<tr>
<td>EOW-S</td>
<td>73.2</td>
<td>25.6</td>
<td>65.4</td>
<td>15.1</td>
<td>-7.8</td>
<td>-0.41</td>
<td>20/32</td>
</tr>
<tr>
<td>CELF-CD-S</td>
<td>3.2</td>
<td>2.2</td>
<td>3.8</td>
<td>1.8</td>
<td>0.7</td>
<td>0.33</td>
<td>24/64</td>
</tr>
<tr>
<td>CELF-RS-S</td>
<td>2.2</td>
<td>1.0</td>
<td>2.3</td>
<td>1.0</td>
<td>0.2</td>
<td>0.11</td>
<td>32/64</td>
</tr>
<tr>
<td>CELF-FS-S</td>
<td>3.7</td>
<td>1.9</td>
<td>2.7</td>
<td>1.2</td>
<td>-1.0</td>
<td>-0.69</td>
<td>56/64</td>
</tr>
<tr>
<td>CELF-CO-S</td>
<td>53.0</td>
<td>10.3</td>
<td>52.7</td>
<td>3.8</td>
<td>-0.3</td>
<td>-0.04</td>
<td>34/64</td>
</tr>
<tr>
<td>RAN-S</td>
<td>3.23</td>
<td>2.73</td>
<td>1.66</td>
<td>0.62</td>
<td>-1.6</td>
<td>0.92</td>
<td>1/16</td>
</tr>
<tr>
<td>NWR-S</td>
<td>69.2</td>
<td>26.7</td>
<td>76.7</td>
<td>7.6</td>
<td>7.5</td>
<td>0.42</td>
<td>19/64</td>
</tr>
</tbody>
</table>

*Note.* See Table 9a note.
Table 10.

*Results of Aligned-Rank procedures comparing three groups on each outcome measure.*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean Rank</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NCP</td>
<td>EL</td>
<td>DEL</td>
<td>λ</td>
<td>AR</td>
<td>p</td>
</tr>
<tr>
<td>CVD</td>
<td>11.7</td>
<td>12.6</td>
<td>11.6</td>
<td>0.005</td>
<td>0.10</td>
<td>0.95</td>
</tr>
<tr>
<td>SSA</td>
<td>9.0</td>
<td>14.0</td>
<td>13.0</td>
<td>0.133</td>
<td>2.79</td>
<td>0.25</td>
</tr>
<tr>
<td>ASM</td>
<td>10.7</td>
<td>15.5</td>
<td>11.6</td>
<td>0.100</td>
<td>2.31</td>
<td>0.55</td>
</tr>
<tr>
<td>ROW-E</td>
<td>14.4</td>
<td>10.8</td>
<td>11.7</td>
<td>0.056</td>
<td>1.30</td>
<td>0.52</td>
</tr>
<tr>
<td>EOW-E</td>
<td>10.4</td>
<td>15.3</td>
<td>12.3</td>
<td>0.091</td>
<td>2.10</td>
<td>0.35</td>
</tr>
<tr>
<td>CELF-CD-E</td>
<td>12.2</td>
<td>15.9</td>
<td>8.5</td>
<td>0.163</td>
<td>3.76</td>
<td>0.15</td>
</tr>
<tr>
<td>CELF-RS-E</td>
<td>8.8</td>
<td>18.3</td>
<td>11.0</td>
<td>0.361</td>
<td>8.30</td>
<td>0.02*</td>
</tr>
<tr>
<td>CELF-FS-E</td>
<td>15.3</td>
<td>12.0</td>
<td>8.7</td>
<td>0.144</td>
<td>3.33</td>
<td>0.19</td>
</tr>
<tr>
<td>CELF-CO-E</td>
<td>10.4</td>
<td>17.9</td>
<td>8.8</td>
<td>0.309</td>
<td>7.12</td>
<td>0.02*</td>
</tr>
<tr>
<td>RAN-E</td>
<td>14.4</td>
<td>11.1</td>
<td>11.2</td>
<td>0.054</td>
<td>1.24</td>
<td>0.54</td>
</tr>
<tr>
<td>NWR-E</td>
<td>8.3</td>
<td>15.3</td>
<td>11.2</td>
<td>0.229</td>
<td>4.82</td>
<td>0.09</td>
</tr>
<tr>
<td>ROW-S</td>
<td>9.4</td>
<td>6.4</td>
<td>9.5</td>
<td>0.090</td>
<td>1.42</td>
<td>0.49</td>
</tr>
<tr>
<td>EOW-S</td>
<td>12.2</td>
<td>14.0</td>
<td>7.8</td>
<td>0.115</td>
<td>2.42</td>
<td>0.30</td>
</tr>
<tr>
<td>CELF-CD-S</td>
<td>11.2</td>
<td>15.5</td>
<td>10.7</td>
<td>0.095</td>
<td>2.18</td>
<td>0.34</td>
</tr>
<tr>
<td>CELF-RS-S</td>
<td>11.5</td>
<td>14.2</td>
<td>12.0</td>
<td>0.031</td>
<td>0.71</td>
<td>0.70</td>
</tr>
<tr>
<td>CELF-FS-S</td>
<td>15.7</td>
<td>12.5</td>
<td>8.8</td>
<td>0.169</td>
<td>3.88</td>
<td>0.14</td>
</tr>
<tr>
<td>CELF-CO-S</td>
<td>12.5</td>
<td>14.0</td>
<td>7.8</td>
<td>0.121</td>
<td>2.67</td>
<td>0.26</td>
</tr>
<tr>
<td>RAN-S</td>
<td>8.9</td>
<td>10.5</td>
<td>7.0</td>
<td>0.073</td>
<td>1.16</td>
<td>0.56</td>
</tr>
<tr>
<td>NWR-S</td>
<td>11.9</td>
<td>15.3</td>
<td>9.8</td>
<td>0.093</td>
<td>2.14</td>
<td>0.34</td>
</tr>
</tbody>
</table>

*Note.* Raw scores from standardized language tests were used for these analyses. The NCP, EL, and DEL columns report the mean rank of the unstandardized residuals, aligned for pretest score and for age, for the group. Higher ranks indicate higher change scores. The \( \lambda \) statistic is a measure of the variance explained by the dependent variable;
for example, for SSA $\lambda = 0.133$ indicates that 13.3% of the variance in SSA change scores is explained by group membership. The AR statistic is calculated from $\lambda$ and follows a chi-squared distribution. $p$ reports the probability of the observed AR statistic using the chi-squared distribution with 2 degrees of freedom. The post-hoc results column reports between-group comparisons for tasks that showed significant omnibus testing results.

* $p < 0.05$
Figures

Figure 1.

Summary of procedures across time.

Figure 1. The order of study procedures across time is represented from top to bottom.
Figure 2.

Percentage improvement across languages for each participant.

*Figure 2.* Legend displays two task categories: English, and Spanish. To calculate percentage of improvement within a category for a participant, each task was scored as +1, 0, or -1 (see Tables 8a, 8b, 8c) and all tasks across a category were summed. The total was then divided by the number of tasks in the category to create a percentage.
Bibliography


attention view of working-memory capacity. *Journal of Experimental
Psychology: General, 130*, 169-183. doi: 10.1037//0096-3445.130.2.169


Kohnert, K. (2004). Cognitive and cognate-based treatments for bilingual aphasia: A

CA: Plural.

Kohnert, K. (2010). Bilingual children with primary language impairment: Issues,

Kohnert, K.J., & Bates, E. (2002). Balancing bilinguals II: Lexical comprehension and
cognitive processing in children learning Spanish and English. *Journal of
Speech, Language, and Hearing Research, 45*, 347-359. doi:10.1044/1092-
4388(2002/027)

semantic production and cognitive processing in children learning Spanish and
English. comprehension and cognitive processing in children learning Spanish
and English. *Journal of Speech, Language, and Hearing Research, 42,* 1400-1413. PMid:10599623


Kohnert, K., Windsor, J., & Pham, G. (2009, June). *Separating differences from disorders using processing-dependent measures*. Poster presentation at the Symposium on Research in Child Language Disorders, Madison, WI.


PMid:9430746


doi:10.1016/S0010-9452(08)70276-4


Warren, S.F., Fey, M.E., & Yoder, P.J. (2007). Differential treatment intensity research: A missing link to creating optimally effective communication


Appendices

Appendix A.1

*Full text of Parent Satisfaction Survey, in English.*

**Parent Satisfaction Survey**

We would like to know more about your experiences participating in the bilingual speech/language program offered by the University of Minnesota (during afterschool). Please indicate your level of agreement with each statement by marking the appropriate box.

<table>
<thead>
<tr>
<th>Agree strongly</th>
<th>Agree</th>
<th>Neither Agree nor Disagree</th>
<th>Disagree</th>
<th>Disagree Strongly</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.  My child receives a high quality service from the University of Minnesota speech/language program.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. The activities and materials used in the program match my child’s age and interests.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. In general, I am happy with the program.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. The clinicians my child works with are interested in helping my child.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. The clinicians make speech/language therapy interesting and fun.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. The program has helped my child communicate better with his/her family and friends.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. I feel that I can ask the project coordinator and clinicians questions.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. I am satisfied with the amount of service my child receives in the program.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Thank you for your time. Please return this paper in your child’s backpack.

Appendix A.2
*Full text of parent satisfaction survey, in Spanish.*

**Cuestionario de Satisfacción de Padres**

Quisiéramos saber más sobre sus experiencias participando en el programa de habla-lenguaje bilingüe ofrecido por la Universidad de Minnesota (durante el programa después del día escolar). Favor marcar la casilla apropiada para indicar su nivel de acuerdo con cada de las siguientes declaraciones.

<table>
<thead>
<tr>
<th>De acuerdo</th>
<th>Completa-mente de acuerdo</th>
<th>No de acuerdo</th>
<th>Completa-mente en desacuerdo</th>
<th>Ni de acuerdo o no de acuerdo (sin opinión)</th>
</tr>
</thead>
</table>

1. Mi hijo(a) recibe servicios de alta calidad por el programa de habla-lenguaje de la Universidad de Minnesota.

2. Las actividades y materiales utilizados en el programa corresponden a los intereses y de acuerdo a la edad de mi hijo(a).

3. En general, estoy contento(a) con el programa.

4. Los patólogas/terapistas quienes trabajan con mi hijo(a) están interesados en ayudarlo.

5. Los patólogas/terapistas hacen divertidas e interesantes las terapias.

6. El programa ha ayudado a mi hijo(a) comunicarse mejor con su familia y sus amigos.

7. Siento que puedo hacer preguntas a la coordinadora del programa y patólogas/terapistas.

8. Estoy satisfecho(a) con la cantidad de servicios que mi hijo(a) recibe en el programa.
Gracias por su tiempo. Favor de devolver este formulario en la mochila de su hijo(a).
Appendix B.

*Full text of Child Satisfaction Survey.*

Child Satisfaction Survey
We want to know what you think about coming to being in our study. For each question mark the face that shows us how you feel.

<table>
<thead>
<tr>
<th>YES</th>
<th>MAYBE</th>
<th>NO</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Smile" /></td>
<td><img src="image" alt="Neutral" /></td>
<td><img src="image" alt="Sad" /></td>
</tr>
<tr>
<td>It was fun to come to speech.</td>
<td><img src="image" alt="Smile" /></td>
<td><img src="image" alt="Neutral" /></td>
</tr>
<tr>
<td>I can listen and talk better because I was in the study.</td>
<td><img src="image" alt="Smile" /></td>
<td><img src="image" alt="Neutral" /></td>
</tr>
<tr>
<td>The games I played weren’t too hard or too easy. They were just right.</td>
<td><img src="image" alt="Smile" /></td>
<td><img src="image" alt="Neutral" /></td>
</tr>
<tr>
<td>My teacher in the study really helped me.</td>
<td><img src="image" alt="Smile" /></td>
<td><img src="image" alt="Neutral" /></td>
</tr>
</tbody>
</table>