The effectiveness of two grammar treatment procedures

for children with SLI: A randomized clinical trial

Karen M. Smith-Lock

Australian Research Council Centre of Excellence in Cognition and its Disorders

Department of Cognitive Science

Macquarie University

Suze Leitão

School of Psychology and Speech Pathology

Curtin University

Polly Prior

Department of Education of Western Australia

Lyndsey Nickels

Australian Research Council Centre of Excellence in Cognition and its Disorders

Department of Cognitive Science

Macquarie University

Author Note

Karen M. Smith-Lock, Australian Research Council Centre of Excellence in Cognition
and its Disorders, Department of Cognitive Science, Macquarie University; Suze Leitao
School of Psychology and Speech Pathology, Curtin University; Polly Prior
Department of Education of Western Australia; Lyndsey Nickels, Australian Research
Council Centre of Excellence in Cognition and its Disorders, Department of Cognitive
Science, Macquarie University. Correspondence concerning this article should be addressed
to Dr. Karen Smith-Lock, ARC Centre of Excellence in Cognition and its Disorders,
Macquarie University, North Ryde NSW Australia 2109; karen.smith-lock@mq.edu.au

Abstract

Purpose: This study compared the effectiveness of two grammar treatment procedures for children with Specific Language Impairment (SLI).

Method: A double blind superiority trial with cluster randomization was used to compare a cueing procedure designed to elicit a correct production following an initial error, to a recasting procedure, which required no further production. Thirty-one 5-year-old children with SLI participated in eight, small group, classroom-based treatment sessions. 14 children received the cueing approach and 17 received the recasting approach.

Results: The cueing group made significantly more progress over the eight week treatment period than the recasting group. There was a medium-large treatment effect in the cueing group and a negligible effect size in the recasting group. The groups did not differ in maintenance of treatment effects eight weeks post treatment. In single subject analyses, 50% of children in the cueing group and 12% in the recasting group showed a significant treatment effect. Half of these children maintained the treatment effect eight weeks later.

Conclusion: Treatment that used a structured cueing hierarchy designed to elicit a correct production following a child's error resulted in significantly greater improvement in expressive grammar than treatment which provided a recast following an error.

Introduction

Children with Specific Language Impairment (SLI) have a well-documented difficulty

acquiring grammatical morphemes such as past tense or third person singular 's' (e.g.,

Leonard, 2000; Rice, Wexler, & Cleave, 1995). This grammatical difficulty can affect the

child's ability to communicate successfully in a range of social and academic contexts. For

example, participating in conversation requires competence with a range of tenses to maintain

the timing or sequence of events in the conversation. Consequently, morpho-syntax has long

been a focus of treatment for children with language impairment (Fey, Long, & Finestack,

2003).

A number of studies have demonstrated treatment efficacy using a variety of techniques

to develop morpho-syntax (e.g., Camarata & Nelson, 1992; Fey, Cleave, Long & Hughes

1993; Leonard, Camarata, Pawlowska, Brown, & Camarata, 2006; Haskill, Tyler & Tolbert,

2001; Tyler, Lewis, Haskill & Tolbert, 2002). Two small-scale studies specifically contrasted

techniques which required a child to produce a grammatical target after an error to those that

did not, and found that the technique requiring production resulted in better treatment

outcomes (Connell & Addison-Stone, 1992; Weismer & Murray-Branch, 1989). On the other

hand, in another small study, a recast conversational approach was shown to require fewer

presentations to generalise to spontaneous speech than a direct imitation approach, although

both were successful (Camarata & Nelson, 1992). It appears that a variety of techniques can

be efficacious, but not all the parameters of efficacious treatment are thoroughly understood.

It is vital to know whether an intervention program that has evidence for its efficacy in

ideal conditions is also effective in everyday clinical practice. To date, evidence of treatment

effectiveness for interventions targeting language skills in school-aged children has been

sparse (Cirrin & Gillam, 2008). Smith-Lock, Leitão, Lambert & Nickels (2013a) and Smith-

Lock et al. (2013b) reported the effectiveness of grammar treatment programs implemented

in classroom settings. They found that grammar treatment delivered to 5-year-olds with SLI, in small groups, in a classroom setting can be effective.

This small body of efficacy and effectiveness research that suggests that expressive morpho-syntax can improve with treatment. These studies have typically contained small numbers of children and have focused on establishing whether or not grammatical impairments can be treated, with few delving systematically into the effectiveness of the various techniques employed. With this existing body of evidence that suggests grammar can improve with treatment, we are now ready to systematically consider the "active ingredients of intervention" (McCauley & Fey, 2006, p.10), that is, the techniques and procedures used during teacher-student interaction.

Several researchers have argued that intervention that is highly effective is that in which the adult response is contingent upon the child's errors (Juel, 1996; Schuele & Boudreau, 2008). As such, the speech-language pathologist or teacher should use feedback to focus the child on the critical aspects of the target, as in the use of a recast, where on producing an error, the child is provided with the target response. The underlying rationale is that hearing a recasted production informs or highlights to the child that their utterance was incorrect. This relies on the child having 1) the metalinguistic awareness to compare his production to the teacher/clinician's, 2) the pragmatic skills to understand he has been corrected, and 3) a focus on the grammatical aspects of the production rather than the semantics, in order to identify the nature of the correction. This is not necessarily as complex as it sounds, as recasting in a classroom often takes the tone of a correction, as that is expected of a teacher-student interaction. Nevertheless, the provision of a correct model following any production, whether correct or incorrect, does not provide the child with explicit feedback. If, however, the recast is provided with some emphasis on the error made (e.g., "He run**sss**," with a longer and/or louder [z] in response to "he run") this might serve to

draw attention to the grammatical target in question. In addition, the provision of a second opportunity to *produce* the target may also communicate to the child that their first attempt was unsuccessful. Thus, if, following an error, the child is encouraged to try again, s/he may understand that his/her first attempt was incorrect and, furthermore, s/he will have another opportunity to produce a correct production.

Vigil & Van Kleeck (1996) (cited by Schuele & Boudreau (2008)) emphasised that the therapist or teacher should respond to a child's error in a manner that would support the child achieving the target or goals. Such scaffolding should include information as to whether or not the response is correct (e.g., recast with emphasis, request for a further attempt), and should be constructed to provide gradually more structure to the child in encouraging a correct response. Thus, if the child fails in his/her second attempt at a grammatical structure, provision of a choice between a correct and incorrect production should provide more help for the child than the recast with emphasis on the grammatical target provided after the first error. The choice provides a correct model, but also contrasts the incorrect and correct productions, potentially increasing the child's awareness of the grammatical structure in question.

Bearing in mind that the goal of the teaching is to facilitate a correct grammatical production, it makes sense that each teaching episode *end* in a correct production, where possible. Thus, if the provision of a choice between correct and incorrect productions still does not result in a correct production by the child, more support is required. At this point, a simple request for imitation (sometimes called a mand) provides even more information to the child regarding the expected response. Thus, for this study, we designed a cueing procedure that used a pre-planned scaffolding hierarchy focusing on the specific target for each child. The procedures used within each teaching episode began with an elicitation within a structured activity and, in the case of an incorrect response, moved gradually

through stages of further support contingent on response (request for clarification, recasting with emphasis on the grammatical target, provision of a forced choice, and elicited imitation), with the goal of completing the teaching episode with the successful production of the grammatical target.

There are a number of components of our approach that are consistent with milieu teaching, a natural conversation-based teaching procedure, including strategies such as: environmental arrangement, selection of specific language targets, the use of specific imitation-modeling prompts, functional reinforcement of the child's communicative attempts and the use of simple recasts (Hancock & Kaiser, 2006). Milieu teaching has been shown to be effective in the earlier stages of language development, in children with autism, intellectual disability and language impairment (Hancock & Kaiser, 2006). However, rather than incidental teaching in naturalistic contexts, we implemented our treatment within small groups in the classroom in structured activities designed around specific targets, commencing with an elicitation of the target (e.g., "Tim, your clown has a hat! Whose hat is it?").

With these treatment principles in mind, we aimed to compare the effectiveness of two pre-planned forms of feedback to a child's error in the treatment of grammar in children with SLI: recasting versus a cueing hierarchy. Both treatments used structured activities designed to model and then elicit specific morpho-syntactic targets followed by adult feedback in response to a child's attempt at the target utterance. Both treatments followed a correct response by modelling the correct target back to the child accompanied by non-specific feedback. The treatment approaches only differed in the procedures followed in response to an error by the child. We predicted that treatment that provided opportunities for the child to produce the grammatical target correctly following an error through the use of a pre-planned scaffolding hierarchy which provided increasing support for the child and ended

in the child producing the target correctly, would be more successful than treatment which provided feedback through the use of recasts without opportunity for production.

We asked the following experimental questions.

1. Did treatment procedure affect treatment outcomes?

2. Did the groups differ in maintained gain eight weeks after treatment?

3. Were group results reflected in individual participant's results?

Method

Design

This was a double blind superiority trial which compared the effectiveness of two treatments for grammar impairment in SLI.[1] Cluster randomization was used, with treatment randomly assigned by site.

Participants

Participants were drawn from a specialized school for children with language impairment, Language Development Centres (LDCs), in Western Australia. 31 students were included in the analysis. The children had a mean age of 5 years, 1 month (61 months). All were monolingual speakers of standard Australian English. SES was determined using the Socio-Economic Index for Areas (SEIFA) Index of Relative Socio-Economic Advantage and Disadvantage Score (RSAD) for the participants' home addresses (Australian Bureau of Statistics, 2011). RSAD rankings ranged from the 24th to the 98th percentile.

Participant flow is illustrated in the Supplementary Materials. This represents the total number of participants at each stage of the project. All 64 children in their first year of full time schooling were invited to participate in the study and assigned to treatment conditions,

---

[1] A superiority trial is designed to compare one treatment to another. Double blind refers to the fact that neither testers nor child participants were aware of their treatment condition, or indeed, that they were participating in a treatment study.

as treatment was part of their regular classroom program. 45 out of 64 students agreed to

participate in our study (see Participant Flow "exclusions: no consent"). Seven of the children

who agreed to participate were deemed ineligible due to diagnoses other than SLI, or to non-

English speaking background, leaving 38 children (see Supplementary Materials: Participant

Flow, Exclusions: Outside clinical criteria). Of these 38 children, five were absent from

school for one of the testing sessions and were therefore excluded from the analysis, leaving

33 children. The timing of their absences can be tracked in Supplementary Materials:

Participant Flow. Absent children were not excluded from subsequent treatment or testing.

Two children did not complete the treatment program, leaving 31 children. One of these

children was distressed by having to change classroom for the treatment sessions; the second

child moved out of the area. Of the 31 children included in the analysis, 17 received the

recasting procedure and 14 children received the cueing procedure. This sample size gave us

the ability to detect an effect size of .4, with power of 0.8. The recasting group contained four

girls and the cueing group, two girls.

Diagnosis of language impairment.

Children were diagnosed with SLI both on entry to the LDC and at the time of this

study. Entry to the LDC required diagnosis as specifically language impaired by a speech-

language pathologist. Children were assessed using the Clinical Evaluation of Language

Function –P2[2] (Wiig, Secord, & Semel, 2006) as one part of an extensive assessment process

for referral to the school. Referral information also included evidence that children had non-

verbal skills in the normal range, as attested by a psychologist or paediatrician. For this

purpose, referring clinicians used a variety of tests, including the Wechsler Preschool and

Primary Scale of Intelligence (WPPSI-III) (Wechsler, 2002), Cognitive Adaptive Test

(Accardo & Capute, 2005), Denver Developmental Screening Test (Frankenburg, Dodds, &

---

[2] One child was assessed with the CELF-P Semel, Wiig, & Secord, 1992

Archer, 1992) and Griffiths Mental Development Scales (Griffiths, 1970). We reviewed the

school's intake data and included only those children who were unequivocally diagnosed as

SLI according the school's intake criteria. We also confirmed the diagnosis of each child with

staff speech-language pathologists (SLPs) at time of testing. Finally, all children were tested

on the Test of Early Grammatical Impairment (TEGI) (Rice & Wexler, 2001), and the

Wechsler Non-Verbal Intelligence Test (WNV) (Wechsler & Naglieri, 2006) at the time of

this study. Only children who failed at least one subtest of the TEGI and who scored above

85 on the WNV were included. Standardized test data can be found in table 1.

All testing and intervention was carried out at the children's schools (LDCs). The study

was approved by the Macquarie University Human Ethics Committee and the Department of

Education, Western Australia.

Insert table 1 about here

Group Allocation

The participating school was located across two sites. Cluster randomization was used

to allocate treatment conditions to the sites, with one site randomly allocated the recasting

procedure and one site randomly allocated the cueing procedure. The two sites were under

the same administration and therefore did not differ in class size or access to professional

development. There is no reason to believe that the general level of teaching expertise or the

language environment differed between the sites. The sites drew on similar socio-economic

populations.

Test Materials

Three tests were used in the project: the Grammar Screening Test, the Articulation

Screening Test, and the Grammar Elicitation Test. These tests were designed specifically to

identify treatment targets and to measure grammatical change and are described in detail in

Smith-Lock et al. (2013a, 2013b). In both the Grammar Tests, children were shown a series

of pictures, and asked a question for each picture designed to elicit a specific grammatical structure. For example, to elicit the regular past tense marker -ed, the child was shown a picture of a boy kicking a ball. The tester then asked "What did this boy do yesterday?". Vocabulary items used were not used in treatment activities.

The purpose of the Grammar Screening Test was to identify potential treatment targets for each child. The test consisted of six possessive *s*, six regular past tense marker -ed, six regular third person singular marker -3s. Two items were included for each allomorph ([s], [z] and [əz]; [t], [d] and [əd]).

The purpose of the Articulation Screening Test was to establish if the children had the articulation skills to produce the targeted grammatical goals. The children were asked to repeat 26 single syllable non-words which contained the final consonant clusters in the grammatical items tested (e.g. pept, avz).

The purpose of the Grammar Elicitation Test was to obtain multiple instances of a particular grammatical target to act as a pre-and post-test measure of treatment effectiveness. The test contained three sections: possessive, regular third person singular marker -3s, regular past tense marker -ed. Each section consisted of 30 items, divided into ten for each allomorph. Each child completed only the section of the test that applied to his or her grammatical target (i.e., 30 items). Items were presented in a different random order for each round of testing. This test has an inter-rater reliability of .99 (Pearson's Product-Moment Correlation, Smith-Lock et al. (2013a, 2013b)).

Each child first completed the Grammar and Articulation Screening Tests. Following the screening tests, a potential treatment target was selected for each child, according to the procedure outlined below. The Grammar Elicitation Test was then carried out, in order to collect further data on each child's selected grammatical target. The Grammar Elicitation Test was administered four times, at eight week intervals. Test 1 and Test 2 provided a pre-

treatment baseline. Treatment occurred between Test 2 and Test 3. Test 3 measured

immediate post-treatment skills. Test 4 measured post-treatment skills eight weeks after

treatment.

A grammatical construction was considered a potential treatment target if the child

made at least six errors on the Grammar Elicitation Test. If a child made errors on several

target constructions, the construction which occurred first on a developmental continuum was

chosen, in the following order: possessives, regular past tense marker -ed, regular third

person singular marker -3s.

Interventions

Treatment consisted of weekly 1 hour sessions for eight weeks. We contrasted the

teacher/SLP's response when the child made an error, across two treatment conditions. For

both conditions, a single treatment session involved both whole class (approximately 12

children) and small group activities. First, the staff SLP delivered a whole group lesson

which incorporated the three grammatical goals. Children were then divided into three

groups, based on their identified treatment target. Each small group carried out three

activities designed to teach them their specific grammatical target. The activities were led by

the SLPs, classroom teachers and education assistants, based in each Language Development

Centre. Teachers/SLPs were provided with detailed activity plans, scripts and vocabulary. A

variety of activities were used, such as dress-up, modelling clay, play with toys, board games

and books. For example, the dress-up activity involved children dressing up as a particular

character. As an item of clothing or a prop was removed from a bag, the group discussed

whose item it was (e.g., the fireman's badge). The child playing the fireman would then get

the badge to wear. The activity with modelling clay involved each child carrying out an

activity with the clay and discussing it. For example, the teacher/SLP rolled her clay and

then said "I rolled the clay." She then instructed the next child, for example, to squish the

clay. The child was expected to carry out the action and then comment "I squished the clay."

The same activities were used in both treatment programs, on the same schedule. Sample

treatment plans can be found in Supplementary Materials: Appendix A. The number of

children in each group ranged from 2 to 5. The entire session, including the whole class

introduction and the three small group activities, lasted one hour.

During the treatment sessions, in both conditions when the child responded correctly,

the teacher/SLP was directed to provide another model of the target (e.g., "he runs"), provide

nonspecific feedback (e.g. "well done") and move on to the next child in the group. The

treatment conditions differed in how the teacher/SLP responded to a child's error. In the

recasting group, (see Supplementary Materials: Treatment Procedure), following an error, the

correct answer was provided to the child, but no attempt was made to have the child produce

the target correctly. In the cueing group, when the child made an error the teacher/SLP

followed a hierarchy of cues designed to elicit a correct answer. These cues provided

progressively more support to the child. If a child answered correctly at any point in the

hierarchy, the teacher/SLP then followed the procedure for a correct response, and provided a

model of the correct target and non-specific feedback. Sample transcripts of teacher/SLP and

child interactions can be seen in Supplementary Materials: Appendix A. All treatment

sessions were audio recorded.

Outside of these treatment sessions, teachers were asked to carry out their classroom

program as they normally would. They were provided with a list of the items that would be

targeted during the treatment. Teachers were not discouraged from modelling or reinforcing

the targets in whatever fashion they normally would throughout the day, but neither were

they specifically asked to do so.

Dosage

Dosage was measured in two ways: 1) number of trials per child and 2) number of models heard. Data were obtained from the transcription and analysis of one full treatment activity (15 minutes per teacher/SLP) during week 7 of the treatment program. Initial plans to determine dosage from records kept during treatment were abandoned after observation of treatment sessions suggested that the records may not be accurate, due to the multiple demands faced by the teachers/SLPs in running the activity, adhering to the treatment protocol and managing behaviour.

Audio recordings were transcribed using an AltoEdge USB foot pedal to allow recordings to be slowed for accurate transcription. Because the recordings were made in a noisy environment, the transcriber flagged any concerns she had regarding the accuracy of the transcription. If these flags affected scoring, the audiotapes were listened to again by the scorers and a decision was reached by consensus. Transcripts were coded first by a research assistant, then a second time jointly by the first two authors. Any disagreement was resolved through discussion and mutual agreement between the second scorers. Two transcripts (13%) were rescored for reliability (one from each condition). Agreement reliability for 'total models heard' was 98%.

Treatment Fidelity

Treatment fidelity was facilitated by a number of measures: staff training, manualised therapy activities, detailed recording of children's responses and observations of treatment sessions. At the beginning of the project, teachers, speech-language pathologists and education assistants attended a full day education session which provided hands-on practice of the treatment activities and techniques relevant to their group allocation. Before each treatment session, teachers/SLPs were provided with detailed documentation for each activity, including scripts and materials and a "cueing card" as a reminder for how to respond to a child's error. Throughout each session, a record was kept of the children's responses to

each item. At least once during the treatment program, each teacher/SLP was observed by the first author. This provided an opportunity for the teacher/SLP to ask specific questions, observe a demonstration of the treatment protocol within their treatment activity and to receive personal feedback regarding their implementation of the treatment.

Treatment fidelity was measured directly in two ways: 1) interviews with school staff and perusal of treatment records to determine structural fidelity and 2) transcription and analysis of treatment sessions to determine adherence to protocol. One activity (15 minutes) for each small group was transcribed and scored, as described above for dosage. The same session (week 7, activity 1) was transcribed for each group so as to analyse all treatment providers with the same experience in the program. This week was selected as it provided the most complete data set, with recordings of the staff who most frequently delivered the treatment. Twelve different activities were represented in the transcribed sample. Agreement reliability for adherence to protocol scoring was 96%.

Blinding

Both testers and participants (children) were blind to treatment conditions. The children involved in the study saw the treatment as a regular part of their classroom activities and had no contact with children in the other treatment condition. The children were accustomed to language instruction and regular testing and therefore can be considered blind to the research process.

Two testers carried out the pre and post intervention testing and a third research assistant scored the tests. Each tester tested the same children in each testing phase to reduce the likelihood of test score changes being due to different testers. The testers and the scorer were blind to the nature of the study. They were told they were studying grammatical development in the children, but were unaware of the treatment component of the project. Post-study interviews confirmed that the blind testers had remained blind to the purpose of

the testing. By necessity, the teachers/SLPs were not blind to the intervention condition they
were administering.

Data Analysis

We used both a standard group comparison approach comparing the experimental

groups, and a case series approach, in which each child acted as his own control in a single-

subject design (Logan, Hickman, Harris, & Heriza, 2008; Nickels, 2002). In each analysis,

the dependent variable was performance on the Grammar Elicitation Test.

Question 1. Did treatment procedure affect treatment outcomes?

To determine if treatment procedure affected treatment performance, we compared the

gain made between Test 1 and Test 2 (pre-treatment gain) with the gain made from Test 2 to

Test 3 (post-treatment gain), for both groups. Pre-treatment gain was calculated by

subtracting Test 1 scores from Test 2 scores. Post-treatment gain was calculated by

subtracting Test 2 scores from Test 3 scores. In order to control for the potential effect of IQ

on treatment outcome, a 2 x 2 mixed analysis of covariance was conducted, with one

between-groups factor (treatment procedure: recasting, cueing) and one repeated measure

(gain: test 2- test 1, test 3 – test 2), with IQ as a covariate. This was followed by planned

comparisons (ANCOVAs) to determine if the groups differed from each other in their gain

scores over the pre-treatment and treatment periods. In the ANCOVAs, pre-test score was

included as a covariant in order to control for the possibility that those children with lower

pre-test scores had greater potential for gain. Thus, test 1 was a covariant in the between

groups comparison of pre-test gain (test 2 - test 1) and test 2 was a covariant in the between

groups comparison of post-test gain (test 3 – test 2).

Question 2: Did the groups differ in maintained gain eight weeks after treatment?

In order to determine if maintenance of treatment gain eight weeks after treatment

concluded differed between the groups, an analysis of covariance was conducted. Treatment

procedure was the independent variable and follow-up gain (test 4 - test 3) was the dependent

variable. Post-test gain (test 3- test 2) was included as a covariate, in order to control for the

fact that the amount of change at follow-up might be affected by treatment gain.

Question 3. Were group results reflected in individual participants results?

Each individual subject's treatment progress was addressed by comparing each child's

gain from to Test 1 to Test 2 (pre- treatment baseline) to their gain from Test 2 to Test 3 (pre-

treatment to post-treatment). For each item, gain was calculated, yielding 30 data points per

child. The gain scores for each child for the two different periods were then compared using

Related Samples Wilcoxon Signed Rank Tests.

Post-treatment maintenance was examined by comparing test 4 performance to test 3

performance, for each child who showed a significant treatment effect, using Weighted

Statistics for Comparison of Levels (WEST-COL) with one sample t-tests (Howard, Best &

Nickels, 2015). A child's post-treatment scores for each item were weighted by minus one

and the delayed post-treatment scores were weighted by plus one. Scores were weighted in

this manner so that if there were no difference between post- and delayed post-tests, the total

across the two tests would be 0 (-1 + 1 = 0). For each participant, a one sample t-test was then

carried out to determine whether, across the 30 items, the sum of the weighted scores differed

significantly from zero, indicating a significant difference in the post and delayed post-test

performance.

All statistical analyses were carried out with SPSS Statistics Software (version 19), with

the exception of effect size, which was calculated using Cohen's *d*, with an online Effect Size

Calculator (http://www.cognitiveflexibility.org/effectsize/) and the WEST-COLs, which were

calculated using Excel.

Results

Baseline Equivalence

Age at time of treatment, standardized language scores and results of statistical

analyses can be found in Table 1. The two groups did not differ significantly in age, or in

standardized language scores. Nonverbal IQ was, however, close to significance and so was

included as a covariate in the analysis. They did not differ significantly on either test 1 or test

2 of the Grammar Elicitation Test (results in table 3 and discussed further below) (test 1:

$t(29) = 1.22$, $p = .23$; test 2: $t(29) = .63$, $p = .53$). In the recasting group, 10 children targeted

possessive, three regular past tense marker -ed and 4 regular third person singular marker -3s.

In the cueing group 8 children targeted possessive, 1 child targeted regular past tense marker

-ed and 5 targeted regular third person singular marker -3s.

Dosage

Total models heard during one, fifteen minute, treatment activity was calculated for

each group. Models were considered any correct production of the grammatical target heard

during the activity. Mean number of models heard in the recasting group was 86.22 (SD =

20.05) and in the cueing group was 87.67 (SD = 52.92). This difference was not statistically

significant ($t(13) = 0.69$, p=.95). If this figure is extrapolated across the whole treatment

program, cumulative intervention intensity for the recasting group was 2069 (86.22 (dose) x 3

per week[4] (dose frequency) x 8 weeks (total duration of intervention)) and for the cueing

group was 2104 (87.67 (dose) x 3 per week[3] (dose frequency) x 8 weeks (total duration of

---

[3] Number of models for the group teaching session is unavailable and has therefore not been

included. This, therefore, underestimates total number of models. Although sessions took

place weekly, three activities took place per session, hence 3 times per week in this

calculation.

intervention)). (See Warren, Fey & Yoder (2007) for a detailed discussion of cumulative intervention intensity).

Number of target utterances produced per child was calculated by dividing the total number of child utterances by the number of children in the group.[4] The mean number of target utterances per child in the recasting group was 10.08 (SD = 3.16) and in the cueing group was 14.29 (SD = 7.05). This difference was not significant (independent groups t-test, two tailed: t(13)=1.57, p = .14).

Treatment Fidelity

Interviews with school staff and running records indicate that all eight sessions were delivered, once weekly, as planned. Consistent with the effectiveness research and real-life conditions in schools, staffing was sometimes different than anticipated, due to staff absences. Only staff who had attended the professional development delivered treatment, with the exception of one teacher who stepped in for one session. In her case, treatment fidelity was fostered through personal discussion with the SLP and the manualised program provided to her. Attendance records indicated that 27 out of 31 children (87%) attended at least seven of the eight treatment sessions. No child missed more than two sessions.

Transcripts were scored to determine the proportion of target utterances that were followed by the correct treatment protocol. Each teacher/SLP response to a child was scored as either adhering to protocol or not. No partial scores were given. Results can be seen in Table 2. Independent t-tests, two-tailed, indicated no significant differences between the groups (proportion of target utterances followed by correct protocol: t(13)=1.17, p = .27; proportion of target errors followed by correct protocol: t(13)=1.52, p = .15; proportion of correct target utterances followed by correct protocol: t(13)=0.15, p=.89).

---

[4] Calculating a direct sum per child from the transcripts was not considered reliable, as it was not always possible to identify the speaker in the recording.

The most common protocol error in the cueing group (70% of protocol errors following a child error) was the provision of one of the cues in the hierarchy, but not the correct cue according to the predetermined hierarchy, whereas the most common error in the recasting group (60% of protocol errors following a child error) was to ignore an error and not respond with a recast, or any feedback at all.

Insert Table 2 about here

Question 1. Did treatment procedure affect treatment outcomes?

Results of the Grammar Elicitation Test can be seen in Table 3. The covariate, IQ, was not significantly related to gain score ($F(1,28) = 0.11$, $p = .75$, partial $\eta^2 = .004$). There was no significant main effect for gain score ($F(2,28) = .287$, $p = .6$), a significant main effect for treatment procedure ($F(2,28) = 7.1$, $p = .01$), and no significant interaction ($F(2,28) = 1.01$, $p = .32$), after accounting for the effects of IQ. Despite the lack of an interaction, the planned comparisons were carried out. The covariate, test 1, was significantly related to gain score from test 1 to test 2 (pre-test baseline) ($F(1, 28) = 4.45$, $p = .04$, partial $\eta2 = .14$). There was no effect of group on gain score from test 1 to test 2, after controlling for the effect of test 1 score ($F(2,28) = 0.09$, $p = .77$, partial $\eta2 = .003$). The relationship between the covariate, test 2 and gain score from test 2 to test 3 approached significance ($F(1, 28) = 3.98$, $p = .06$, partial $\eta2 = .124$). There was a significant difference between the groups for Test 2-Test 3 gain scores ($F(2,28) = 4.32$, $p = .047$), after controlling for the effect of test 2 score. Thus, the groups were making similar progress in the eight weeks prior to intervention, but the cueing group made significantly more progress over the eight week treatment period than did the recasting group.

The cueing group showed an effect size of .74 from Test 1 to Test 2 (pre-treatment baseline) and 1.49 from Test 2 to Test 3, considered a large treatment effect (Cohen, 1988). In contrast, the recasting group showed an effect size of .68 between Test 1 to Test 2 (pre-treatment baseline) and .85 between Test 2 and Test 3. Thus, there was a medium-large treatment effect in the cueing group when compared to the effect size over the pre-treatment baseline (1.49 - .74 = .75), whereas the recasting group showed a negligible effect size (0.85 - .68 = .16).

Insert table 3 about here

Question 2: Did the groups differ in maintained gain eight weeks after treatment?

Analysis of post-treatment maintenance indicated that the covariate, post-test gain (test 3-test 2) was significantly related to post-treatment maintenance (test 4 - test 3) ($F(1, 28) = 20.15$, $p = .001$, partial $\eta2 = .42$). There was no significant difference between the groups for Test 4-Test 3 gain scores, after controlling for treatment gain ($F(2,28) = 0.5$, $p = .83$, partial $\eta2 = .002$). Thus, the groups did not differ significantly in post-test maintenance after controlling for treatment gain.

Question 3. Were group results reflected in individual participant's results?

Results for each child can be seen in Figures 1 and 2. Seven out of the 14 children (50%) in the cueing group and two out of 17 children (12%) in the recasting group showed significantly greater gain over the treatment period from Test 2 to Test 3 than in the pre-treatment baseline period from Test 1 to Test 2. Of the nine children who demonstrated significantly greater gain over the treatment period than in the pre-treatment baseline period, four children (57%) in the cueing group and one child (50%) in the recasting group showed no significant difference between performance in the post-test (test 3) and performance in the delayed post-test (test 4), indicative of maintenance of the treatment effect.

No trends were identified in demographic or diagnostic variables that might explain the differential responses to treatment amongst the children.

<div align="center">Insert figures 1 and 2 about here</div>

Post hoc Question: Was treatment success related to grammatical target?

Because children were allocated a treatment target based on individual profile, it was possible that the groups differed in the number of children targeting each grammatical item, and therefore in the difficulty of the treated targets. Since it has been argued that verbal inflections are more difficult for children with SLI to acquire than nominal inflections (e.g. Rice, Wexler & Hershberger, 1998), performance on the two inflection types was compared. The two groups had similar distributions of nominal and verbal treatment targets. Treatment gains were made in both nominal and verbal inflections, with no effect of target in either group (cueing: Fisher exact test (1-tailed) p = .41; recasting: Fisher exact test (1-tailed) p = .68). It appears that treatment success cannot be attributed to treatment target.

As noted above, seven children were excluded from the analysis; five because they missed a testing session, one because he moved out of the area and one because she was distressed at changing classrooms for the treatment. Of these seven children, two were from the cueing group and five from the recasting group. These seven children did not differ from the main group on the basis of age (mean age of absent children: 58.43 months (SD = 4.6 months; mean age of those included: 60.76 months SD= 3.45, t(43) = 1.56, p = .13), or intake language score (mean CELF-P2 Core Language score of absent children: 71.71 months, SD = 9.48 months; mean CELF-P2 Core Language score of those included: 66.63 months SD = 18.61, t(43) = 0.66, p = .52).

<div align="center">**Discussion**</div>

This treatment effectiveness study compared two different procedures for teaching morpho-syntax to 5 year old children with SLI: 'cueing' and 'recasting'. The cueing

condition, which consisted of a pre-planned cueing hierarchy designed to end in a correct production of the target following an initial error, led to a greater treatment effect than the condition which provided a recast after an error (i.e., without requiring production). Group analyses indicated a large effect size for the cueing group, but not for the recasting group. The groups did not differ significantly in the maintenance of treatment effects after eight weeks. Single subject analyses showed that half of the children showed significant gain following the cueing treatment, whereas few children benefitted from the recasting treatment. Half of the children in each group who made a significant gain in performance maintained that gain after eight weeks. Treatment success was not related to which grammatical target was treated. These results suggest that the choice of treatment techniques and procedures does make a difference to treatment outcome.

The results of this effectiveness study are consistent with most, but not all previous efficacy studies. For example, Fey, Cleave, Long & Hughes (1993) and Leonard, Camarata, Pawlowska, Brown & Camarata (2006) reported success with focused stimulation and conversational recasts in grammar treatment for three to five year olds. Our study did not find such success with a similar procedure (the 'recasting' approach), although some children did improve in this condition. However, the treatment programs did differ in several key ways. First, in Fey, Cleave, Long & Hughes' study (1993) each individual treatment session with a clinician (received by half of the children) began with direct imitation of the grammatical target and a contrastive target. They did find, nevertheless, that even the group that received no imitation in their treatment (i.e., the group that received parent-administered treatment) showed improvement. Furthermore, their treatment consisted of up to three hours per week for 20 weeks, and targeted four goals using a cyclical approach. Thus their treatment differed from the current study both in frequency (three 1 hour sessions weekly vs one 1 hour session), duration (twenty weeks versus eight weeks) and number of targets (4 vs 1).

Similarly, the study by Leonard, Camarata, Pawlowska, Brown & Camarata (2006) involved 96 treatment sessions, far more than the current study. These differences suggest that recasting treatment may be effective over longer treatment periods. In addition to the treatment differences, the measure of treatment success in Fey et al's study was the Developmental Sentence Score, a measure of grammatical skill in spontaneous speech rather than use of a specific morpheme in a structured task as in our study.

The results of the current study are consistent, however, with the findings of Connell and Addison-Stone (1992) and Weismer and Murray-Branch (1989), both of which directly compared modelling techniques with techniques that involved imitation or evoked production. The treatment duration of both of these studies was much shorter than those of Fey, Cleave, Long & Hughes (1993) and Leonard, Camarata, Pawlowska, Brown and Camarata (2006), and more consistent with the study reported here. Thus, the findings of Connell and Addison-Stone (1992) using invented morphemes, and Weismer and Murray-Branch (1989) based on lab-based treatment with four children have been replicated here with a larger group of children in a naturalistic context, adding to the evidence base for treatment effectiveness for morpho-syntax using these techniques.

Our results are also consistent with the findings of Tyler, Lewis, Haskill & Tolbert (2002) whose program targeted children of a similar age over 12 weeks duration (compared to our eight), and resulted in a large treatment effect. The active ingredients of their program included modelling and recasting as well as actively encouraging the children to produce the target through elicited production, similar to those of the present study.

The results for the individual children in our study demonstrated that 50% of those in the cueing group showed significant change, in contrast to 18% of those in the recasting group. This is consistent with Weismer & Murray-Branch (1989) who also used single subject analyses and found that not all children improved significantly. Weismer and Murray-

Branch noted that their unsuccessful participant demonstrated lower receptive language skills. Our study was not designed to test this directly; nevertheless, in examining the data we were unable to identify any demographic or diagnostic factors that explained differential responses to treatment.

Another difference between our study and those discussed above is in measurement of treatment effectiveness. Camarata and Nelson (1992) define 'strong' production as evidence of use of the target outside of the training condition without a direct model and involving untrained words or stems, and 'weak' production being evidence of an elicited target in response to a training stimulus such as a picture with a model using a trained item. Some efficacy studies have measured outcomes by tracking the progress of targets within an analysed a language sample, thus also providing 'strong' evidence (e.g., Fey, Cleave, Long, & Hughes, 1993). We measured outcomes using our Grammar Elicitation Test which, although a structured task, required use of the target outside of the training condition, without a direct model, using vocabulary items not included in the treatment activities. Although we did not measure use of the grammatical target in spontaneous speech, our outcomes show clear evidence of generalization beyond the trained vocabulary items and activities, thus strengthening the evidence for treatment success in our study[5]. Given the time limitations of an in-school effectiveness study, the collection and analysis of in depth language samples was not feasible, however should be added to future studies, if possible.

Let us turn now to the key ingredients of the treatment procedures that might have resulted in the differing outcomes. Both techniques involved intensive modelling of the grammatical target, and we have established that they did not differ in this regard. Teachers/SLPs in both groups responded in the same manner to a correct utterance, that is,

---

[5] For in depth discussion of the theoretical implications of this generalization of treatment, see Smith-Lock (2014).

with a model of the grammatical target as affirmation of the correctness of the child's

response. The key difference between the treatments was the planned response to an error. As

noted in the introduction, several researchers have argued that intervention that is highly

effective is that in which the adult response is contingent upon the child's errors (Juel, 1996,

Schuele & Boudreau, 2008). Certainly the cueing condition differentiated more clearly

between the child's correct responses and his incorrect ones than the recasting condition.

After a correct response, in both treatment groups, the teacher/SLP responded with the

production of a further correct model and general positive feedback, (e.g., "well done").

Following an error, in the recasting condition, the teacher/SLP responded with a simple recast

containing the correct target. Thus, in the recasting condition, the response to the child was

always the correct grammatical target. In contrast, the response to an error in the cueing

technique was the use of an extensive cueing hierarchy working through the pre-planned

steps until a correct production was achieved, a much longer and more involved response

than the recast.

While it is important that the adult response indicate to the child whether his response

was correct or incorrect, it is also critical that the feedback clearly indicate to the child the

source of the error, or indeed, of the success. Once again, this was clearer in the cueing

condition where the teacher/SLP questioned the response by working through the stages of

requesting clarification, repeating the error back to the child, a forced choice question and

finally a recast with a request for imitation (mand). Thus, the cueing provided more specific

feedback to the child that it was the grammar of the response that was incorrect.

Another difference between the techniques was the child's production of a correct

utterance after an error in the cueing condition. Weismer and Murray-Branch (1989)

suggested that production of the target provides a child with the opportunity to practice

production of the morpheme and hence reinforce memory of the correct production.

Similarly, Connell and Addison-Stone (1992) suggested that the requirement to imitate the morpheme resulted in refinements both to the underlying phonological representation of the word as well as provide practice for the child in accessing and producing the motor program.

In sum, the cueing condition provided clear differentiation in the feedback between correct and incorrect responses, clearer information to the child that it was grammatical form which dictated correctness of the response, and more practice in accessing and producing the grammatical representation and motor program of the response. It can be hypothesised that these factors played a part in the greater success of this treatment procedure.

There was no significant difference between the groups in treatment maintenance eight weeks later, once treatment gain was taken into account. In each group, half of the individuals who showed a significant gain in treatment maintained that gain eight weeks later. Thus, treatment procedure did not affect treatment maintenance.

Our treatment program contained no formal maintenance component. Nevertheless, for half of the children, the treatment program as it stands was effective beyond the treatment period. The fact that half of the children did not maintain gains suggests that some children require either a longer duration of treatment or a formal program of treatment maintenance. It is valuable to note that the substantial training and experience that the teachers/SLPs in this study obtained were insufficient in and of themselves to facilitate maintenance for all of the children once the structured classroom programs were withdrawn.

Limitations and Future research

Effectiveness studies, by their nature, do not allow the same extent of control as lab studies. This study took place in two locations. These locations were under shared administration. Staff were provided with the same supervision and professional development. There is no reason to believe the schools differed systematically in staff expertise or experience. The schools were in the same geographic area and drew on similar populations.

The groups did not differ on expressive or receptive language score. However, comprehension of the specific grammatical targets was not measured specifically prior to treatment. While the targets were expressive ones, children with comprehension difficulties are often those who respond least well in much of the intervention reported in the research (Ebbels, 2014). Comprehension of the grammatical targets may well be a factor that affects an individual's treatment success, and thus should be measured in future studies.

Furthermore, we were limited in our ability to collect and analyse language sample data for evidence of generalization from structured tasks to spontaneous speech. While the use of our Grammar Elicitation Test provides strong evidence of generalization in a structured task, language sample data would further strengthen the findings.

Our analysis of treatment transcripts found no significant differences in dosage or in adherence to the treatment protocol in the two treatment conditions, however, these results were based on only one activity out of 24 carried out. While it would be preferable to transcribe and analyse a greater percentage of treatment, this in itself was a substantial undertaking, given the nature of effectiveness research. Treatment sessions were delivered in a classroom setting, with three groups carried out simultaneously in one classroom. The recordings were clear enough to be transcribed, but required substantial time and effort. The fidelity analysis involved 17 staff and 12 different activities and as such, can be considered a varied, representative, sample of the treatment. Nevertheless, the transcript findings are based on a relatively small proportion of the overall data and should be treated with caution.

The adherence to treatment, while acceptable, would benefit from improvement. The procedures in the study to facilitate improvement were extensive: full day training, manualised procedure with on-the-desk reminder cards, regular consultation with a speech-language pathologist in the classroom and a personal observation and feedback session with one of the researchers. Nevertheless, it was clear that these techniques were hard to learn and

implement in a busy classroom. Feedback from the teachers/clinicians suggested that the personal demonstration and feedback provided when the researcher joined a teacher's session, was very beneficial. Often the teachers knew what they wanted to do, but struggled to actually make it happen. With the researcher in the session, she was able to intervene precisely when help was needed and demonstrate for the teacher within the teaching episode. It is likely that, in many settings, this intensive, personal instruction will not always be practical to implement. However, it could be augmented by video or audio recording of sessions for the teacher to listen to and evaluate with a mentor. Evaluation of the recordings would require not just that the teacher observe an error they might have made, but also that techniques be provided to avoid that type of error in the future.

It is interesting to note that the nature of the adherence to protocol errors differed somewhat between the two treatment techniques. As stated above, the majority of adherence errors in the cueing group was the provision of an incorrect cue in the hierarchy following an error whereas in the recasting group, lack of adherence was typically the provision of no feedback to the child following an error. Thus, the children in the cuing group still received informative feedback under such circumstances, whereas the children in the recasting group did not. A more in depth analysis of the transcripts, plus the consideration of a larger number of session transcripts, would allow us to explore this difference further.

This study demonstrated that the cueing hierarchy technique was more effective than the recasting technique. It is important to note, however, that it did not evaluate which components of the cueing hierarchy are necessary. It is possible that a single cue to imitate would have resulted in similar outcomes, or that the order of cues in the hierarchy itself is less important than the use of a cue in general.

Further research should focus on further refinement of the key characteristics of treatment success. For example, which components of the cueing hierarchy are necessary; is

it the production component of the cueing technique that drives its success, the cueing hierarchy, or both? Which elements can be removed or streamlined and which are key components to success? Is it essential to maintain the sequence of steps in the cueing hierarchy? Such a detailed investigation can be carried out through controlled experimental manipulation as well as qualitative analysis of treatment transcripts. A second outstanding question is which factors affect an individual's response to treatment. Clearly, linguistic factors such as proficiency prior to treatment or oral language comprehension might play a role, as might non-linguistic factors such as engagement in the treatment process. It is likely that such research will require larger numbers of participants than were available for this project. Thirdly, it remains to be determined whether the treatment procedures differ in the generalization of treatment gains to less structured tasks such as conversation. Finally, the issue of treatment maintenance requires further exploration. Optimal treatment duration, optimal treatment maintenance procedures as well as which factors predict that an individual will maintain treatment gains remain to be determined.

Conclusion

In summary, we found that a grammar treatment program that used a structured cueing hierarchy designed to elicit a correct production following a child's error resulted in significantly greater improvement in expressive grammar than a similar treatment program which provided a recast following an error.

Acknowledgements

References

Accardo, P.J. & Capute, A.J. (2005). *Cognitive Adaptive Test*. Baltimore, MD: Brookes.

Australian Bureau of Statistics, (2011). Retrieved from

http://www.abs.gov.au/ausstats/abs@.nsf/mf/2033.0.55.001

Camarata, S. M. & Nelson, K. E. (1992). Treatment efficiency as a function of target

selection in the remediation of child language disorders. *Clinical Linguistics &*

Phonetics, 6, 167-178.

Camarata, S. M. & Nelson, K. E. (2006). Conversational recast intervention with preschool

and older children. In R. J. McCauley & M.E. Fey (Eds.), *Treatment of Language*

Disorders in Children. (pp. 237-264). Baltimore, MD: Paul H. Brookes Publishing

Co.

Camarata, S. M., Nelson, K.E., & Camarata, M. N. (1994). Comparison of conversational-

recasting and imitative procedures for training grammatical structures in children with

specific language impairment. *Journal of Speech & Hearing Research*, 37, 1414-

1423.

Cirrin, F. M. and Gillam, R. B., 2008, Language intervention practices for school-age

children with spoken language disorders: A systematic review. *Language, Speech,*

and Hearing Services in Schools, 39 (Suppl.), S110–S137.

Connell, P. & Addison-Stone, C. (1992). Morpheme learning of children with specific

language impairment under controlled instructional conditions. *Journal of Speech and*

Hearing Research, 35, 844-852.

Ebbels, S. (2014). Effectiveness of intervention for grammar in school-aged children with

primary language impairments: A review of the evidence. *Child Language Teaching*

and Therapy, *30*, 7-40.

Fey, M. E., Cleave, P. L., Long, S. H. & Hughes, D. L. (1993). Two Approaches to the

Facilitation of Grammar in Children with Language Impairment: An experimental

evaluation. *Journal of Speech, Language and Hearing Research*, 36, 141-157.

Fey, M., Long, S. & Finestack, L. (2003). Ten principles of grammar facilitation for children

with specific language impairments. *American Journal of Speech-Language*

Pathology, 12, 3-15.

Frankenburg, W.K., Dodds, J., Archer, P. et al. (1992). The Denver II: A major revision and

restandardization of the Denver Developmental Screening Test. *Pediatrics*, (89), 91-

97.

Griffiths, R. (1970). *The abilities of young children*. High Wycome, Great Britain:

Cournswood House.

Hancock, T. B. & Kaiser, A. P. (2006). Enhanced milieu teaching. In R. J. McCauley & M.E.

Fey (Eds.), *Treatment of Language Disorders in Children*. (pp. 203-236). Baltimore,

MD: Paul H. Brookes Publishing Co

Haskill, A., Tyler, A., & Tolbert, L. C. (2001). *Months of morphemes*. Eau Claire, WI:

Thinking Publications.

Howard, D., Best, W., & Nickels, L. (2015). Optimising the design of intervention studies:

Critiques and ways forward. *Aphasiology*, 29(5), 526-562.

doi:10.1080/02687038.2014.985884

Juel, C. (1996). What makes literacy tutoring effective? *Reading Research Quarterly*, 31,

268–269.

Leonard, L. B. (2000). *Children with Specific Language Impairment*. Cambridge: The MIT

Press.

Leonard, L. B., Camarata, S., Pawlowska, M., Brown, B., & Camarata, M. N. (2006). Tense

and agreement morphemes in the speech of children with specific language

impairment during intervention: Phase II. *Journal of Speech, Language, and Hearing Research*, *49*, 749-770.

McCauley, R. & Fey, M. (Eds) (2006) *Treatment of language disorders in children*. Paul H. Brookes: Baltimore.

Nickels, L.A. (2002). Therapy for naming disorders: Revisiting, revising and reviewing. *Aphasiology*, 16, 935-980.

Rice, L. & Wexler, K. (2001). *Test of Early Grammatical Impairment*. United States: The Psychological Corporation: A Harcourt Assessment Company.

Rice, M. L., Wexler, K., and Cleave, P. L. (1995). Specific language impairment as a period of exended optional infinitive. *Journal of Speech and Hearing Research*, 38, 850-863.

Rice, M.L., Wexler, K. & Hershberger, S. (1998). Tense over time: The longitudinal course of tense acquisition in children with specific language impairment. *Journal of Speech and Hearing Research*, 41, 1412-1431.

Schuele, C. & Boudreau, D. (2008). Phonological awareness intervention: beyond the basics. *Language, Speech & Hearing Services in Schools*. 39, 3-20.

Smith-Lock, K.M. (2014). Rule-based acquisition of regular past tense in children with specific language impairment. *Cognitive Neuropsychology*. DOI: 10.1080/02643294.2014.951610

Smith-Lock, K.M., Leitão, S., Lambert, L. & Nickels, L. (2013a) Effective Intervention for Expressive Grammar in Children with Specific Language Impairment. *International Journal of Language & Communication Disorders*, 48 (3), 265-282.

Smith-Lock, K.M., Leitão, S., Lambert, L., Prior, P., Cronje, J., Newhouse, S. & Nickels, L. (2013b) Daily or Weekly? The effect of treatment frequency on grammar treatment for children with SLI. *International Journal of Speech-Language Pathology*. 15 (3),

255-267.

Tyler, A. A., Lewis, K. E., Haskill, A., & Tolbert, L. C. (2002). Efficacy and cross-domain effects of a phonology and morpho-syntax intervention. *Language, Speech, and Hearing Services in Schools, 33*, 52– 66.

Vigil, A., & van Kleeck, A. (1996). Clinical language teaching: Theories and principles to guide our responses when children miss our language targets. In M. Smith & J. Damico (Eds.), *Childhood language disorders* (pp. 64–96). New York: Thieme.

Warren, S. F., Fey, M. E. & Yoder, P. J. (2007). Differential treatment intensity research: A missing link to creating optimally effective communication interventions. *Mental Retardation and Developmental Disabilities*, 13, 70-77

Wechsler, D. (2002). *Wechsler Preschool and Primary Scale of Intelligence™ – Third Edition (WPPSI™ – III)*. Sydney, NSW: Pearson.

Wechsler, D. & Naglieri, J. A. (2006). *Wechsler Nonverbal Scale of Ability*. San Antonio TX: Pearson.

Weismer, S., & Murray-Branch, J. (1989). Modeling versus modeling plus evoked production training: A comparision of two language intervention methods. *Journal of Speech and Hearing Disorders, 54*, 269-281.

Wiig, E., Secord, W., & Semel, E. (2006). *Clinical Evaluation of Language Fundamentals Preschool - Second Edition, Australian and New Zealand Standardised Edition* (CELF P-2 Australian and New Zealand). Sydney, NSW: Pearson.

Table 1.

Language standard scores on intake to Language Development Centre (LDC); Age, IQ and language standard scores prior to treatment.

Mean (SD)	CELF-P2[6]			Age[7] in months	WNV[8]	TEGI[9]			
	RLI[10]	ELI[11]	CLS			3rd singular marker -s	Past tense	Be	Do
Recasting Group	70.59 (12.13)	66.24 (20.228)	71.0 (10.7)	60.65 (3.26)	119.06 (8.47)	3 (17.65%)	3 (17.65%)	3 (17.65%)	4 (23.53%)
Cueing	76.43 (13.426)	72.50 (9.957)	73.85 (10.52)	60.78 (3.51)	111.14 (8.47)	2 (14.29%)	0	0	1 (7.14%)
Independent						n/a	n/a	n/a	n/a

group t-tests (2-tailed)	t(29)=2.27 p = .21	t(28)=1.06 p = .3	t(28)=0.73 p = .47	t(29)=0.12 p = .91	t(29)=1.99 p = .06				

[6] Clinical Evaluation of Language Fundamentals - Preschool 2 (Wiig, Secord, & Semel, 2006).

[7] Mean scores, standard deviation in brackets.

[8] Wechsler Nonverbal Test of Intelligence (Wechsler & Naglieri, 2006), Mean scores, standard deviation in brackets.

[9] Test of Early Grammatical Impairment (Rice & Wexler, 2001). Number of participants who passed criterion; percentage of group passed in brackets.

[10] ELI: Expressive Language Index, RLI: Receptive Language Index, CLS: Core Language Score

[11] ELI was unavailable for one child.

Table 2. Adherence to protocol, mean scores, standard deviation in brackets.

Procedure	Total target utterances followed by correct protocol	Target errors followed by correct protocol	Correct target utterances followed by correct protocol
Recasting	78.1% (15.46)	72.07% (20.21)	78.5 % (16.86)
Cueing	66.19% (24.39)	51.22% (33.37)	76.91% (25.06)

Table 3.

Mean score (standard deviation in brackets) on Grammar Elicitation Test (Smith-Lock et al., 2013).

Group	Test 1	Test 2	Test 3	Test 4	Gain Test 1 to Test 2	Gain Test 2 to Test 3	Gain Test 3 to Test 4
Recasting	9.88	13.24	18.59	18.47	3.35	5.35	-.12
	(6.98)	(6.59)	(9.3)	(8.42)	(4.96)	(6.8)	(7.56)
Cueing	6.71	11.57	22.79	17.93	4.86	11.21	-4.86
	(7.39)	(8.08)	(6.9)	(9.37)	(6.39)	(7.94)	(8.72)

Figure Captions

Figure 1. Individual results over 4 tests; cueing group

*significant difference between pre-test gain (test 2-test1) and post-test gain (test 3- test 2)
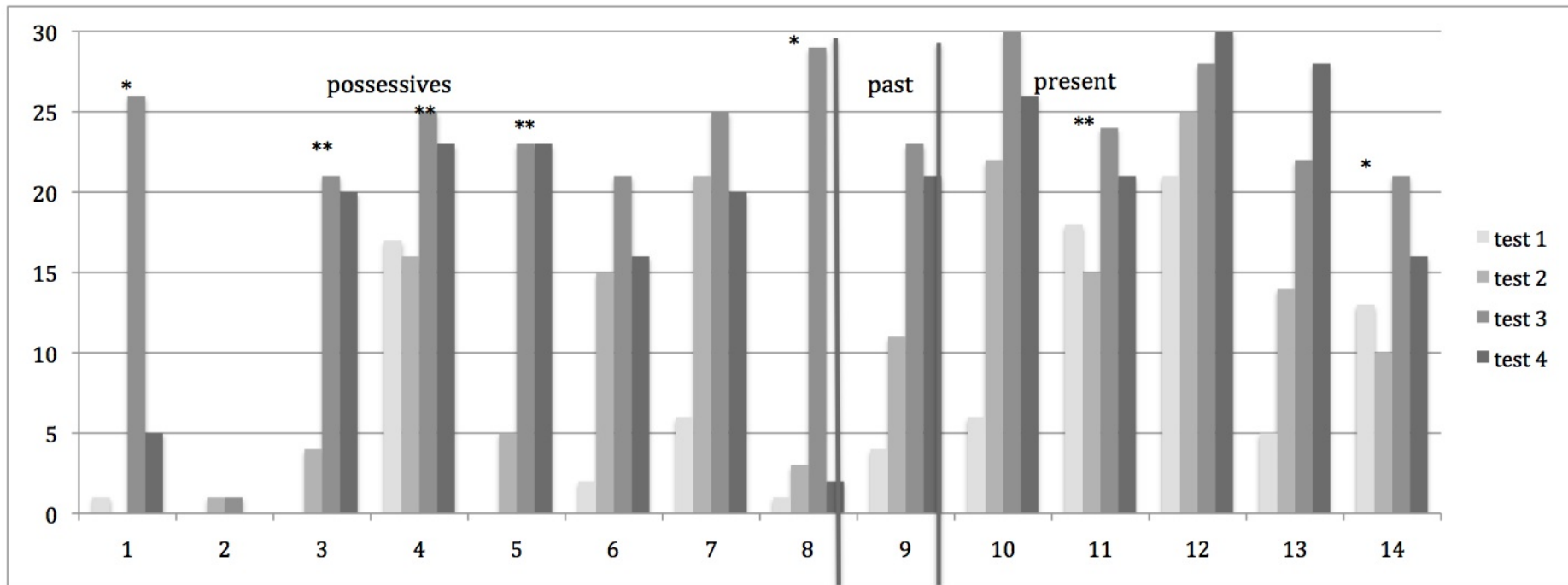
** significant difference between pre-test gain (test 2-test1) and post-test gain (test 3- test 2) and no significant difference between post-test and

delayed post-test

Figure 2. Individual results over 4 tests; recasting group

* significant difference between pre-test gain (test 2-test1) and post-test gain (test 3- test 2)

** significant difference between pre-test gain (test 2-test1) and post-test gain (test 3- test 2) and no significant difference between post-test and
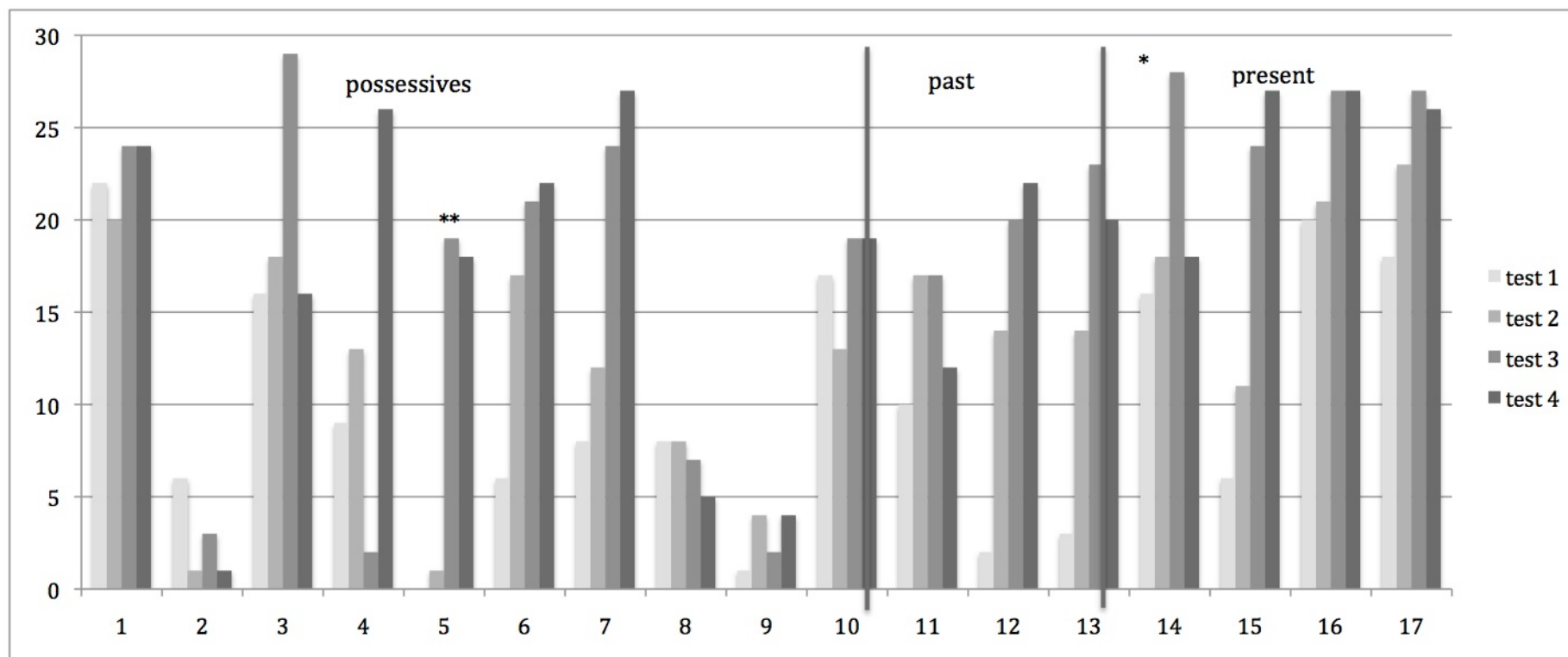
delayed post-test

Figure 1. Individual results over 4 tests; cueing group

*significant difference between pre-test gain (test 2-test1) and post-test gain (test 3- test 2)

** significant difference between pre-test gain (test 2-test1) and post-test gain (test 3- test 2) and no significant difference between post-test and delayed post-test

Figure 2. Individual results over 4 tests; recasting group

* significant difference between pre-test gain (test 2-test1) and post-test gain (test 3- test 2)

** significant difference between pre-test gain (test 2-test1) and post-test gain (test 3- test 2) and no significant difference between post-test and

delayed post-test