



Published in final edited form as:

J Speech Lang Hear Res. 2011 August ; 54(4): 1170–1181. doi:10.1044/1092-4388(2010/09-0246).

Initial MLU Predicts the Relative Efficacy of Two Grammatical Treatments in Preschoolers With Specific Language Impairments

Paul J Yoder,
Vanderbilt University

Dennis Molfese, and
University of Louisville

Elizabeth Gardner
Vanderbilt University

Abstract

Purpose—We sought to confirm predictions based on past findings that pre-treatment mean length of utterance (MLU) would predict which of two grammatical treatments would best facilitate generalized and maintained grammatical development in preschoolers with specific language impairment (SLI).

Method—The participants were 57 preschoolers with specific language impairment (SLI). A randomized group experiment was used. The two grammatical treatments were Broad Target Recasts (BTR) and Milieu Language Teaching (MLT). MLU was assessed at Time 1 in two conversational language samples. Growth rate of productive grammar was quantified using growth curve modeling on the Index of Productive Syntax (IPSyn) from two conversational language samples at each of 6 measurement periods.

Results—Predictions were confirmed for children with initially low MLU, but not for children with initially high MLUs. MLT facilitated growth of grammar better than BTR in children who were initially in Brown's stage I. Effects maintained 5 months after treatment ended.

Keywords

recasts; milieu language teaching; intervention; therapy methods; moderated treatment effects

Compared to their age-mates, many children with SLI are delayed in every major aspect of language. Compared to language-matched peers, children with SLI have special difficulty with grammar (Leonard, 1998). Therefore, the relative efficacy of grammatical treatments is of great interest to clinicians responsible for treating these children.

Correspondence concerning this article should be addressed to Paul Yoder, Peabody Box 228, Vanderbilt University, Nashville, TN 37203. paul.yoder@vanderbilt.edu.

The content of this article is solely the responsibility of the authors and does not necessarily reflect the official views of NIDCD or the National Institute of Health.

The research was conducted by Paul Yoder, Special Education Department, Vanderbilt University; Dennis Molfese, Department of Molecular, Cellular & Craniofacial Biology, University of Louisville; Elizabeth Gardner, Kennedy Center, Vanderbilt University

Child-centered Grammatical Interventions Hold Promise

Child-centered language interventions are correlated with larger effect sizes than therapist-directed ones in producing generalized gains in children's grammar in children with SLI (Yoder & McDuffie, 2002). Child-centered treatments allow the child to select from available materials and determine how to play with them. These interventions more closely resemble the interactions that children experience in the natural environment than do therapist-centered interventions. The similarity between natural conversations and child-centered therapy sessions is thought to aid generalization (Warren & Rogers-Warren, 1985). Additionally, child-centered treatments tend to teach grammatical targets using exemplar utterances that match the child's communicative intent. Presumably, this focus facilitates the integration of semantic and pragmatic aspects of the communicative interaction with new grammatical information provided by the clinician. Such integration is thought to facilitate children in becoming generative with new language skills (Bock, 1982).

Milieu Language Teaching (MLT, Warren, 1991) employs child centered play and routines to elicit the child's need to use utterances that contain pre-selected grammatical targets. The child's current focus of attention or communicative request provides the context and timing for the adult to prompt a child utterance that contains a language structure that the child does not use often or consistently. For example, assume the child does not use auxiliaries or present progressives (e.g., "He is going to the store"). Further assume the child is making an action figure fly. The MLT therapist might say, "What is he doing?" If the child does not respond, the therapist might say, "Say, 'He is flying'". The consequence for a correct child response is often praise (e.g., "Great!") or a recast of the child's utterance to a greater level of complexity. If the child responds, but omits the target structure (e.g., "He fly"), the adult has the option of recasting the child's utterance in a more mature form (i.e., "He is flying") or continuing the child's topic without recasts (e.g., "He's high in the sky"). After all appropriate child responses, compliance to child requests and continuation of the play routine is delivered. There have a great deal of studies, both at the single subject and group level that support the general efficacy of MLT approaches. In the interest of brevity, readers are referred to reviews of MLT to read about the specifics of these studies (Hancock & Kaiser, 2006; Kaiser, Yoder, & Keetz, 1992). In a general sense, these studies as focused on populations other than specific language impairment.

Broad-target Recasts (BTR, Yoder, Camarata, & Gardner, 2005) is another child-led language treatment. BTR uses (a) following the child's play lead, (b) talking about the child's focus of attention, (c) asking the child questions about what he is doing, and (d) grammatically recasting the child's utterances. A grammatical recast repeats a major element of the child's preceding utterance and adds semantic and grammatical information. For example, if the child does not use 3rd person singular, and says, "She sleep", the staff-implemented grammatical recast might be, "Yes, she sleeps". An effective recast is thought to be one that is developmentally progressive of the child's productive language level. There is no emphasis on getting the child to produce a particular exemplar utterance of specific grammatical targets (Camarata, Nelson, & Camarata, 1994). Therefore, the types of questions asked are not designed to get the child to use his cutting edge language skills but are designed to elicit non-imitated child utterances that can then be immediately recasted by the clinician. A recastable utterance is a child intelligible utterance containing linguistic information in utterances other than yes/no, a greeting or an acknowledgement. By "broad target recasts," we mean that target structures in the recast may be any developmentally progressive grammatical structure. The BRT approach has been referred to by several different labels in the literature (e.g., Responsive Intervention, Yoder, et.al., 1995) and is used in many treatment packages (e.g., Hanen It Takes Two to Talk, Pepper & Weitzman, 2004). For example, Yoder, Spruytenburg, Edwards, and Davies (1995) found that broad-

target grammatical recasts of any non-imitative child utterance facilitated generalized increases in mean length of utterance in children with intellectual disabilities. The effect sizes were quite large.

A direct comparison between target-specific and broad target recasting has never been conducted. However, it can be argued that broad target recasting allows more child control over activities and more targets to be addressed than pre-selecting targets and setting up activities that afford platform utterances that afford recasts with the targeted structures. Child control over the activities and the number and type of grammatical structures that are recasted in the BTR approach may allow the child to more actively construct his grammatical knowledge while mapping exemplar sentences to topics the child cares about. Such conditions may contribute to generalizing newly acquired grammatical structures to unfamiliar contexts and exemplar sentences (Yoder & McDuffie, 2002).

There has been a recommendation that the rate of recasting to children with SLI be relatively high compared to that used by parents of typically developing children (Proctor-Williams, Fey, & Loeb, 2001). Using a stage-level target may prove to be one way to enable an accelerated recast rate. Finally, a stage-level goal does not require the same degree of clinician skill as a structure-specific goals approach. It was found to be effective, a greater number of interventionists may be able to use it compared to structure-specific goal selection approaches.

In summary, MLT and BTR are both child-centered treatments. They both use recasts as a consequence to child utterances. MLT uses structure-level grammatical goals selected prior to sessions; whereas BTR uses stage-level grammatical goals selected on the basis of platform utterances provided in the moment. MLT seeks to elicit immediate production of models of the target, and neither MLT nor BTR require imitation of the recast. Both procedures depend on child processing of recasts, though it can be argued that MLT is more likely to prime attention to the target of the recast.

Past Studies Show Relative Efficacy of MLT vs. BTR Varies by Children's Language Level

The broad target approach to recasting was selected, in part, because we have past studies comparing it to MLT and thus a basis for predicting which treatment should be superior for children of differing developmental levels. Two past studies have compared MLT to BTR (Kaiser & Yoder, 1992; Yoder, et al., 1995). The former was a randomized group experiment and the latter was a matched-group comparison design (i.e., a quasi-experiment). In these reports, BTR was called "responsive interaction" but the primary active ingredient was considered to be BTRs. Although the Yoder et al report spoke about "targets," these were only at the broadest level (i.e., single word, 2-word, and multiword expansions). The population in both studies was preschoolers with language delays for various reasons. In both studies, it is unclear what proportion of the samples met the diagnostic criteria for SLI. The pretreatment variables that described the subgroup in which one treatment was more effective for facilitating language outcomes that contain expressive grammatical items or are measures of expressive grammar were (a) receptive and expressive vocabulary age equivalency, (b) receptive and expressive general language age equivalency, and (c) mean length of utterance (MLU). The mean age equivalency across measures and studies for the subgroup that benefited most from MLT was 20 months and below. The mean age equivalency across measures and studies for the subgroup that benefited most from BTR was 34 months and above. MLU was chosen as a putative moderator for the current study because of its specificity and conceptual link to the outcome of high interest in studies of SLI: grammar (Scarborough, Rescorla, Tager-Flusberg, Fowler, & Sudhalter, 1991).

The grammatical targets for MLT are explicitly linked to the children's MLU through the use of Miller's (1991) Assigning Structural Stage to select targets, which presents grammatical structures by MLU stage. As mentioned earlier, one primary difference between MLT and BTR is MLT's use of elicited production prompts for utterances displaying structure-level grammatical targets. Such prompts may be effective in helping children take notice of, and learn from, the prompts and models *if* the child is able to understand the response demand (i.e., what is expected from the child), comprehend the meaning of the prompt or model, and, at least covertly, imitate the model. Another difference between the MLT and BTR treatments is the specificity of preselected targets. Pre-selection of targets are likely to result in more teaching episodes per target, while stage-level targets may allow for input for a greater number of grammatical structures.

It may be easier for children to answer the prompts with the desired grammatical structure for early-learned grammatical structures (e.g., verb+ing) than for later-learned grammatical structures (e.g., 3rd person singular). For example, children may be able to meet the pragmatic requirements of the prompts for complex grammatical structures without using the targeted grammatical structure than for prompts for simpler grammatical structures. Additionally, children may have more difficulty imitating the typically-longer elicited imitation prompts for complex structures than the typically-shorter elicited imitation prompts for simpler structures due to auditory memory constraints.

If any or all of these explanations operate, then the primary processable form of linguistic input in MLT for later-learned targets (and developmentally older children) may be recasts. Because BTR recasts are likely to occur more often than the more complex and target-constrained MLT teaching episodes, there may be more frequent recasts in BTR than teaching episodes in MLT. Thus, for complex targets, BTR may be a more effective treatment than MLT. However, we acknowledge the number of teaching episodes per grammatical structure may not vary in predictable ways between the treatment. This makes the rationale for the predicted greater effectiveness for BTR vs. MLT for children with relatively high MLUs less compelling than the rationale for the prediction for children with relatively low MLUs.

There is an understandable historical mistrust of studies that seek to identify moderated treatment effects by using multiple significance tests without alpha adjustment because the findings of such studies are less likely to replicate than are those of confirmatory studies (Cronbach & Snow, 1977). Indeed, it is only through replicated findings, particularly when the replication is a confirmatory study, that we are able to provide clinically useful information regarding the selection of competing treatments for particular subgroups of children. This study seeks to provide the needed confirmation that language level, in the form of MLU, predicts relative response to these two treatments. This study is the first to address the question in children with SLI.

Finally, there is very little evidence that treatments for children with specific language impairment maintain their gains from language intervention. In their review of the language intervention literature for this population, Yoder and McDuffie (2002) found that that longest follow-up period was only 3 months. Maintenance of gains due to language intervention is important because such is one of the primary characteristics of development in contrast to context-bound behavior change (Yoder & Lieberman, 2008). Therefore, the current study also examines maintenance of grammatical gains.

Research Question and Secondary Purpose

The primary research question is “Does MLU at the pretreatment period predict which of two language treatments is superior in increasing rate of grammatical development in

preschoolers with SLI?” We tested a MLU \times Treatment Group interaction predicting grammatical growth with a prediction that MLT would be superior to BTR for children with relatively low MLU and that BTR would be superior to MLT for children with relatively high MLU. Secondly, we examined the extent to which post-treatment grammatical levels are maintained (or increased) 4 months after the treatment phase ended.

Method

Overview—Children with SLI were tested at approximately 2-month intervals, Time 1 thru Time 6. Measures administered at Time 1 included language testing, nonverbal cognitive procedures, demographic questionnaires, speech accuracy test, two 20-minute language samples, and a sound field hearing screening. Measures administered at Time 2 thru Time 6 included two 20-minute language samples. The children with SLI, the language sample procedures and variables, treatments, and fidelity of treatment procedures are described here.

Participants

Inclusion criteria—Inclusion criteria were set to select preschoolers with SLI who would be appropriate for the treatments. At study entry, children (a) had to live in monolingual, English-speaking homes, (b) have chronological age of 30 mos to 60 mos, (c) have speech accuracy above the 16th percentile as assessed by the Goldman-Fristoe Test of Articulation (GFTA-2, Goldman & Fristoe, 2000) or speech intelligibility of better than 71% (i.e., not more than 1.29 SD below the mean expected at their ages using Chapman & Miller’s, 1981 norms) in a 20 minute language sample, (d) have nonverbal intelligence of at least 83 as assessed by the Leiter International Performance Scale-Revised (Leiter-R, Roid & Miller, 1997; *M* for US = 100, *SD* = 15) and (e) pass a sound-field hearing screening at 30 dB HL at 500, 1000, and 2000 Hz. If the children were adopted from another country (only 1 was), we required that they be over 36 months old, have been in the US for over 2 years and were adopted at or prior to 12 months of age. All children had (a) to have delayed expressive language as evidenced by MLU below 1.29 SD below that expected for their age or score below the 11th percentile on the Expressive Communication Subscale on Preschool Language Scale III (Zimmerman, et al., 1992), and (b) had to fall below the autism cut-off score of 30 on the Childhood Autism Rating Scale (Schopler, Reichler, & Renner, 1999). To increase the probability that the children with SLI were appropriate for both of the interventions, the children’s language during a 20-minute language sample had to demonstrate (a) an average initial mean length of utterance in morphemes (MLU) between 1.0 – 3.5, (b) a total of at least 10 different words, and (c) at least 20 recastable utterances.

Recruitment—The primary method of recruitment was through the directors of local preschools via phone calls and flyers, speech/language screening at local Head Start preschools, and providers of state early intervention service. Parents of participants were informed that assessments and treatments would take place at the university clinic. The payment schedule for assessments at Time 1 was \$50 and at Time 6 was \$100. After parents contacted the research project director, phone screening and scheduling occurred. All recruitment methods, consenting protocols and incentive payment schedules were approved by relevant institutional review boards.

Number and characteristics of participants—Based on power analyses conducted prior to the study, the intended SLI sample size was sixty. The recruited SLI sample size was sixty-two. Fifty-seven preschoolers with SLI received treatment. The two randomly assigned SLI treatment groups were very similar in the putative moderator (MLU), the Time 1 measure of the outcome (IPSyn total score), vocabulary diversity, age, global expressive

language impairment, and nonverbal IQ, (see Table 1). Additionally, the groups were similar on parental formal education level and occupational status. Additionally, all but one of the parents had a high school degree with forty percent of the parents having at least a college degree. The mean occupational status was 53 ($SD = 20$), compared to the national average of 29 ($SD = 23$). Therefore, a large majority of the families were from middle to upper middle socio-economic levels (Stevens & Cho, 1986). Only 11% (6) of the children were under 36 months. The flow of participants is illustrated in Figure 1.

Design

The primary research question was addressed through a randomized group experiment with a longitudinal correlational component. To assign children to treatment groups, a computer program using a random number generator produced the random sequence. Even numbers were assigned to MLT and odd numbers were assigned to BTR. Participants were assigned to numbers in the order in which they were enrolled. The project director enrolled participants and was blind to treatment assignment at the time of participant enrollment. Blocking or stratification on MLU was not conducted prior to random assignment because children with SLI enrolled in the study at a rate of about 1.6 participants per month for the first 3 years of the project. Six measurement points were used to model growth of grammar (study entry, 3.1 mos [$SD = .65$], 5.1 mos [$SD = .66$], 7.1 mos, [$SD = .67$], 9.1 [$SD = .7$], and 11.1 mos [$SD = .68$] from study entry). The fourth measurement period was a post-treatment assessment and the last two were maintenance (i.e., follow-up) assessments.

Differential attrition analysis—Overall attrition was 8% (i.e., 5/62 participants). There was more attrition in the MLT group (4/28) than in the BTR group (1/34), $X^2(1) = 3.8, p = .051$. However, there was no statistical interaction between drop-out status and treatment group predicting the Time 1 measure of the outcome variable (IPSyn total score), even using a liberal .20 alpha due to low power owing to the small drop-out sample size. Additionally, of the 21 pretreatment variables on which we had scores for the drop outs (including demographic variables, standardized language scores, IQ score, and Time 1 IPSyn scores), two were significantly different between the completers and the drop-outs using the liberal .20 alpha: IQ ($t(55) = 1.75, p = .09$, favoring the drop-outs) and occupational status ($t(55) = -1.73, p = .09$, favoring the completers). Neither of these pretreatment variables was associated with either intercept or slope of the IPSyn growth curves nor did they statistically interact with treatment group to predict either growth parameter of the IPSyn.

Between-treatment differences in non-project treatment participation—The children with SLI participated in nearly zero non-project language intervention. The means for number of hours per month during the treatment phase were 0.78 ($SD = 1.4$), and 0.67 ($SD = 1.8$) for BTR and MLT, respectively. These means were not different between treatment groups, $t(55) = 1.00; p = .32$.

Procedures

Language samples—This procedure was used to derive measures of pretreatment MLU (Time 1) and measures of grammatical growth rate over the entire period study involvement (Times 1 thru 6). At all time periods, two 20-minute language samples were conducted. Two samples were used to maximize the probability of generating the number of utterances needed for the measure of grammar (i.e., 100). To reduce measurement error, the examiner used a well-specified interaction style with all children and prescribed sets of toys, rotated by time period on an established schedule to counter boredom. To assess far-transfer of trained grammatical structures, the examiner, toys, and interaction style differed from those used in treatment sessions.

Orthographic transcription and coding of language samples: All language samples were video recorded, compressed, and digitized. ProCoderDV software (ProCoderDV, Tapp, 2003) was used to guide control of the media file during orthographing and coding of all samples. All orthographing and coding was implemented by staff specifically trained for this duty. All child utterances from the language samples were orthographed and coded for morphemes as prescribed in Systematic Analysis of Language Transcripts (SALT, Miller & Chapman, 1993). The orthographed and coded language sample pairs were concatenated to yield one larger sample for analysis. The mean length of utterance (MLU) in morphemes was derived from the complete and intelligible utterances in the concatenated language sample. At all measurement periods, the concatenated language samples were grammatically analyzed using the Index of Productive Syntax (IPSyn, Scarborough, 1990) module of the Computerized Profiling program (CP, Long, Fey, & Channell, 2000). After CP coding of the language sample, each file was reviewed for accuracy and edited by a trained speech-language pathologist (SLP). The IPSyn has been found to be a valid indicator of grammatical level and reflects the use of structures listed on the same taxonomy used to generate grammatical goals and guide selection of recasting (i.e., Miller's, 1991, Assigning Structural Stage Charts). The IPSyn score reflects the child's use of particular major syntactic and morphological structures in the language sample (Scarborough, 1990)

Project treatments—Children with SLI were randomly assigned to either Milieu Language Teaching (MLT) or Broad Target Recasting (BTR). Therapy sessions were offered on a schedule of 30-minute sessions, 3 sessions per week, for a 6-month period in a 1:1 format at a university clinic. Clinicians were certified SLPs who were trained for 1 month in their prescribed treatment prior to implementing it with participants. Quality and consistency of treatment implementation was monitored by a senior-level SLP, trained by the first author in the treatments.

MLT: MLT therapists used environmental arrangement, a system of graduated prompts for children to produce targeted language structures, and recasts, functional rewards, or verbal rewards to facilitate child language. From a language sample, three language targets were selected on the basis of absence of utterances exemplifying a structure from a list of potential targets that are typically present in children with the target child's MLU. The list of targets is indicated in Miller's (1991) Assigning Structural Stage charts. These targets were replaced when children used three nonimitative examples of the structure in treatment sessions. The context of treatment was child-directed play. For example, if a child in the one word stage of language development says "Ball" while reaching for a ball to roll down a ball maze, the adult may withhold the ball while looking expectantly at the child. If the child does not respond, the adult may ask, "What do you want?". If the child still does not respond while attending to the ball or adult, the adult may say, "Say, 'Want ball'" (i.e., elicited imitation request). If the child says, "Want ball", the adult might say, "You want the ball" and give the child the ball. Only the level of prompt needed to elicit the desired response was used and faded as soon as possible.

BTR: BTR clinician used environmental arrangement and prompts to elicit child production of recastable "platform" utterances, and then responded with grammatical recasts to facilitate grammar. The context of treatment was child-directed play. The BTR clinician recasted any recastable child utterance that afforded modeling a developmentally-appropriate grammatical structure. Developmental appropriate recasts (a) retained the meaning of the child's platform utterance, (b) added grammatical information to the child's platform utterance, and (c) contained a grammatical structure that fell into, or one stage above, the child's current MLU stage based on Miller's Assigning Structural Stage (Miller, 1991). Although questions were used to elicit child talk, they were not used to elicit child

production of the language targets. Elicited imitation requests were never used in BTR. For example, if the child was in the one word stage of language and was playing with cow from the farm set, the adult might have asked the child what the cow is doing. If the child said, “cow eat”, the adult might say, “The cow is eating”. Thus, BTR includes many of the key elements seen in other reports for recast interventions (e.g., Camarata, Nelson & Camarata, 1994), but also differs in two aspects: (a) BTR does not require target pre-selected grammatical structures and (b) prompts to elicit children’s platform utterances are included in BTR.

Fidelity of treatment: To document the intensity and quality of treatment implementation, trained coders orthographed and coded all child and adult utterances from the middle 15 minutes of 3 treatment sessions that were recorded during the 2nd, 4th, and 6th months of the treatment phase for each participant. The coder was never the clinician implementing the treatment. Children’s attendance to scheduled treatment sessions was carefully recorded. To quantify intensity of exposure to presumed active ingredients, the rate per minute of correctly-implemented MLT episodes and the rate of BTR grammatical recasts were derived. To quantify the quality of MLT implementation, the proportion of MLT teaching episodes that were correctly implemented were derived from the coded fidelity of treatment sessions filmed. A correct MLT teaching episode began with an adult prompt intended to elicit from the child one of the preselected, targeted grammatical structures and ended with an adult production of the targeted structure in the form of either a recast of the child’s response or a model of the target. To quantify the consistency of BTR implementation, the proportion of recast opportunities actually recasted by the clinician was calculated. An opportunity for adult recasting occurred when a child utterance included at least one intelligible word that functioned in a grammatical role. An adult recast was coded when it (a) followed a child utterance that contained at least one intelligible, grammatical word within 2 seconds, (b) incorporated at least one grammatical word that from the child’s utterance, (c) adhered closely to the child’s topic, meaning, and intent, and (d) modeled a developmentally appropriate grammatical target. In addition to the filming and coding of periodic treatment sessions, every 2 weeks, one session for each child in treatment was observed by a staff member who tallied and reported the number of MLT teaching episodes or BTR recasts to the clinician, providing timely feedback

Attendance to sessions was approximately 96% (BTR $M = 68.9$, $SD = 3.1$; MLT $M = 68.6$, $SD = 3.2$) of offered (i.e., 72) sessions and was not different between treatment groups, $t(55) = -.22$, $p = .8$. The rate of the presumed active ingredient (grammatical recast for BTR and correct teaching episode for MLT) per minute was greater for BTR ($M = 4.3$, $SD = .74$) than for MLT ($M = 3.2$, $SD = .57$), $t(55) = 5.6$, $p < .001$, $d = 1.65$. The average proportion of teaching episodes that were correctly implemented in the MLT group was .89 ($SD = .17$). The average proportion of opportunities recasted in the BTR group was .80 ($SD = .07$).

Non-project treatment description—To allow description and analysis of whether non-project treatment participation affected the results, parents completed a form at each of the 6 time periods, reporting the average number of hours of language therapy their child had received outside of their participation in the project. This information was derived for the total number of participants, for each treatment group, and for individual participants.

Interobserver Reliability

A random sample of 20% of the coded data was independently recoded to estimate the interobserver reliability of the coded variables. Primary observers were blind to which sessions would be checked for reliability. Reliability for language samples was derived from (a) independently generated transcriptions, orthographed by a second transcriber and (b)

independent coding of these secondary transcriptions. Intra-class correlation (ICC) coefficients using the absolute agreement criterion were used to estimate interobserver reliability because among-participant variability in the true score of variables influences the degree of interobserver reliability that is required to provide a reasonable test of research questions and because ICC reflects both the among-participant variance in true score estimates and measurement error due to disagreements between-observers (Cronbach, Gleser, Norda, & Rajaratnam 1972). The ICC for MLU at Time 1 was .98. The ICCs for the edited IPSyn total score for Times 1–6 averaged .86 ($SD = .07$, range = .76 – .95). The ICC for the rate of recasts per minute averaged .97 ($SD = .01$, range = .96 – .98). The ICC for rate of correct teaching episodes per minute was .95 ($SD = .02$, range = .94 – .98).

Because one expects nearly perfect implementation from an efficacy study that is carefully monitored, nearly zero variability among participants in (a) the proportion of MLT teaching episodes that are correct and (b) the proportion of opportunities that are recasted was expected. Therefore, percentage agreement was a more reasonable estimate of interobserver reliability for these two variables than was ICC (Yoder & Symons, 2010). Average percentage agreement on the proportion of MLT teaching episodes that were correct and the proportion of opportunities that were recasted was .91 ($SD = .11$) and .97 ($SD = .02$), respectively.

Results

Preliminary Results

Mixed level modeling was used to quantify rate of growth because it enables an efficient and more informative way to quantify growth than traditional repeated measures end-point analyses such as repeated measures ANOVA or MANOVA (Singer & Willett, 2003). Table 2 indicates the descriptive statistics for the IPSyn total score at each of the 6 assessment periods. The values for Time variable were set to “0” at study entry, while Time values at other assessment periods were time from entry in months. This allowed treating Time as a continuous variable and allowed interpreting the intercept of growth curves as an estimate of initial IPSyn total score. The Level 1 model indicated that a straight line fit the change in IPSyn total scores. The average growth curve, regardless of treatment group was as follows:

$$\text{Estimated IPSyn total score} = 38.97 + 1.49(\text{time in study in months})$$

This equation means that the average IPSyn total score at study entry was about 39 points and the average growth was about 1.5 points for each month in the study. Both the intercept, $t(56) = 23$, $p < .001$, and slope, $t(56) = 12$, $p < .001$, were significantly different from zero (i.e., had significant fixed effects). Additionally, the intercept, $X^2(56) = 324$; $p < .001$, and slope, $X^2(56) = 87$, $p < .001$, had significant variance (i.e., had significant random effects).

Primary Analyses Relevant to Research Questions

Does pretreatment mean length of utterance (MLU) predict which of two grammatical treatments is superior in facilitating growth in grammar in preschoolers with SLI?—We used mixed level modeling (Bryk & Raudenbush, 1992) in which the Level 2 variables were treatment group (coded $-.5$ and $.5$, as suggested by Kraemer, Kiernan, Essex, & Kupfer, 2008), grand-mean-centered MLU at Time 1, and the product term of MLU x Treatment. We found that the fixed effect of the product term predicted slope of the growth in IPSyn total score, $t(53) = -2.01$, $p = 0.049$. Other statistics from the Level 2 model are in Table 3. For example, there were no main effects of treatment group on growth of grammar. The amount of explainable variance accounted for in the slope was .57. Using the method indicated by Preacher, Curran, and Bauer (2006), the lower point

along the MLU continuum at which the groups' growth curves differed between-groups was MLU 1.84 and under. This method is an empirical method for identifying the cut-off point(s) on the moderator at which the between group differences on the outcome are significant. The upper region of significance was not interpretable. An illustration of the interaction is provided in Figure 2.

To what extent was there maintenance of grammatical production level after post-treatment?—The proportion of the SLI sample that maintained or gained in the IPSyn total score from post-treatment to the 4-month follow up was 74% (77% in the BTR, 67% in the MLT) and was not different between groups. In fact, there was significant growth in the IPSyn total score from Time 4 to Time 6, $t(56) = 4.0$, $p < .001$, $d = .44$.

Do the between-group differences at Time 1 explain variance in grammatical growth rate?—We tested whether the variables in Table 1 that showed the largest (albeit nonsignificant) between-group difference at Time 1 (IQ and formal education) were related to the growth in the IPSyn. They were not. The p values for the association of IQ and formal education with slope of growth in the IPSyn was .32 and .65, respectively.

Discussion

The predictions were partially confirmed. Within the SLI sample, the average rate of growth in the measure of grammar, the IPSyn total score, was about 1.5 points per month in the study, but there was, of course, much variability in the rate of grammatical growth. Children who began treatment with MLUs at or below 1.84 showed faster grammatical development when treated by MLT than when treated by BTR. There were no between-treatment group differences for children who began treatment with MLUs above 1.84. Finally, a large majority of the participants maintained their grammatical growth 4 months after treatment and, as a group, the sample showed a moderate gain in grammatical development between the post-treatment and 4 month follow-up periods.

As was suggested in the introduction, it is hypothesized here that the developmental level of the targeted grammatical structures influenced MLT's relatively greater effect in comparison to BTR for children with relatively low MLUs. It may be easier for children to respond with the expected grammatical structure to the MLT prompt and to imitate the model for early-learned grammatical structures than for later-learned grammatical structures. Additionally, children may have more difficulty imitating the typically-longer elicited imitation prompts for complex structures than the typically-shorter elicited imitation prompts for simpler structures. The added requirement to produce phrases and sentences with the targeted structure may primary children's attention to that structure. Unfortunately, we did not have the resources to conduct follow-up analyses to garner evidence that might support these explanations. Another explanation might be that children who are just beginning grammar development may need the greater specificity of the MLT's structure-level goals compared to the stage-level goals of BTR.

As with any study in which predictions are not confirmed, our explanations are more tenuous. There are multiple differences between the current study versus extant studies that supported a prediction that BTR would be superior to MLT for developmentally older children. Two primary differences offer a feasible explanation that could be the topic of future study: population and outcome. The extant studies addressed language as global outcomes in children with developmental delays. The current study focused on grammar in children with specific language impairment. Children in the current study who began treatment with MLUs above 1.84 may have been struggling with grammatical targets that are particularly resistant to change (e.g., 3rd person singular). Such goals may need multiple

teaching episodes directed at the structure (Leonard, Camarata, Pawlowska, Brown, & Camarata, 2008).

The finding that MLU moderates relative treatment response confirms a very specific prediction made prior to the study. In fact, the current study's finding that MLT is superior to BTR in facilitating grammatical development in children who are developmentally young is remarkably similar to that found in the only two studies that compare MLT and BTR (Kaiser & Yoder, 1992; Yoder et al. 1995). These studies found that the average age equivalency below which MLT was superior to BTR was 20 months. Using the regression equation for the association between MLU and chronological age in typically developing preschoolers and toddlers, the MLU expected at 20 months is 1.51 (Miller & Chapman, 1981). In the current study, the empirically-derived value along the MLU continuum at which the slope of growth in grammar was significantly greater in MLT than in BTR (i.e., 1.84) was in the same language stage as 1.51 (i.e., Brown's late stage I, Brown, 1973). This level of convergence on both moderator and cut-point for the region of significance is unprecedented in the language intervention literature.

Convergence among studies examining the moderators of differential effect of two treatments serves to support the value of both the exploratory studies that set the stage for the prediction and for the value of the current confirmatory study. Such replicated confirmatory studies are needed to identify (a) which characteristics of children we should attend to (i.e., the moderator) and (b) the cut-off value of the moderator we should use to select a treatment for a particular child (Brown's stage I).

Knowing that most children that begin treatment in Brown's stage I acquire grammar faster in MLT than in BTR provides a basis for initial decision making without other specific information about the child's possible response to the treatment. Without such information, clinicians may select a particular treatment based on personal preference, familiarity, or a misplaced belief in its superior effectiveness for all language impaired preschool age children.

Limitations

The current study would have been stronger if it had used (a) blind coders and examiners, and (b) more fidelity of treatment assessments. Although, the lack of blindness can cause between-group differences where there are actually none in the population (i.e., type I errors), the complex findings of the current study are very unlikely to be due to systematic over- or underestimation of children's true grammatical level depending on the coder's knowledge of the child's initial MLU level and group assignment. Although, more fidelity of treatment assessments would provide a more stable estimate of the level of adherence to intended treatments than is available in the current study, the lowered confidence in the accuracy of our estimate of the true level of treatment adherence during the treatment phase does not detract from the primary findings of the study. Additionally, observational coding of accuracy and amount of active ingredient use is arguably superior to descriptions based on the more commonly-used checklist approach to measuring treatment fidelity.

Including grammatical growth patterns for children who were assigned to treatment but did not finish the treatment in the same analysis with children who did complete treatment (i.e., an intent-to-treat analysis) would have improved the accuracy of the effect size estimate for children to whom treatment is offered (Frangakis & Rubin, 1999). However, the present study showed little attrition (8%) and there were small and nonsignificant between-treatment group differences on all measured pre-treatment variables, including the moderator and Time 1 measure of the outcome. These facts mitigate our concerns regarding the potential

negative consequences of using an “analysis of the treated” approach to testing the research questions instead of an “intent to treat” approach.

In the current study, as in almost all studies of moderated treatment effects, the moderator (MLU) was only a correlate of the differential treatment effect on grammatical development. This finding is subject to the same weakness of all correlational findings. That is, there may be other, perhaps unmeasured, variables correlating with MLU (e.g., age, number of teaching episodes per grammatical structure) that are the causal influence on the differential treatment effect. Although this issue cannot be completely ruled out using research designs that do not manipulate the putative moderator, there are two reasons to consider MLU a very good proxy for the causal influence on the superior effect of MLT in developmentally young children. First, it is unlikely that the association of MLU and differential treatment effect is sample-specific because the cut-off value and the moderated treatment effect in the present study replicates results found in two previous studies (Kaiser & Yoder, 1992, Yoder et al., 1995). Second, our explanation for this finding specifically links MLU and the differences between the two treatments.

Finally, this study tested whether MLU predicted the relative efficacy of BTR and MLT. The results do not inform was regarding the relative efficacy of MLT vs. target-specific recasting, or other treatments not tested.

Strengths

Numerous attributes support the inference that this study produced trustworthy, clinically-useful results. Rigorous random assignment procedures were used to allocate participants to treatments. Many tests of pre-treatment differences confirmed pre-treatment equivalence on many putative correlates of the outcome. Fidelity of treatment data verified that both treatments were carried out with a high degree of adherence to the protocols. The rate of recasts in the BTR treatment, which was partially motivated by a desire to deliver a high rate of recasts, exceeded the recommended level for children with SLI (Proctor-Williams, 2001). The interventionists were trained professionals who were monitored regularly by a master clinician. Differential attrition and non-project treatment participation were unlikely alternative explanations for the between-treatment group differences on the outcome. The sample was average to large for treatment studies on children with SLI (Law, Garrett, & Nye, 2004; Yoder & McDuffie, 2002). The measure of grammar came from an ecologically-valid measurement context for preschoolers: conversational language samples. These language samples constituted tests of far-transfer as multiple dimensions of the measurement context were different from the treatment contexts (i.e., person, activities, materials, interaction style). For example, prompts for child talk and recasts were not used during the language samples. There was evidence of maintenance of the grammatical gains four - five months after the end of the treatment phase. Finally, the primary findings of the study replicated findings from the extant literature at a very specific level of analysis.

Importance of the findings

Pre-treatment MLU can be used to guide initial treatment selection. However, results of group research designs are only beginning points for selecting treatments for an individual. There are individual differences within the subgroups described by the moderator variable and cut-point for the region of significance. Therefore a clinician would be wise to use these study findings to inform the choice of treatment for a particular child, monitor progress over a period of a few months, and shift to another treatment if the individual child does not show progress. For example, imagine a scenario with a skillful MLT clinician whose teaching episodes are properly timed to the child’s teachable moments. This clinician determines that the choice of MLT therapy is an appropriate treatment choice based on analysis of the

child's language sample. She identifies appropriate grammatical goals in Brown's late stage I, and begins regularly scheduled treatment sessions with the child, periodically monitoring for progress. After several months, the child, who appears to resist elicited production prompts, has not shown progress in grammatical development. At this point, the clinician should reduce or eliminate elicited production prompts and emphasize the recast consequences for the child's utterances.

Acknowledgments

The research was funded by NIDCD R01DC005994 and received support from the Vanderbilt Kennedy Center for Research on Human Development (P30HD15052). Much gratitude is expressed to Allison Cody-Swope, Noelle Hardin, Abigail Kirsch, Nicole Thompson, Rebecca Abel, Cher Savas, and Mary Camarata for their capable assessment and clinical work. Thanks are also extended to Stephen Camarata, who provided helpful comments on the paper. Mostly we are grateful to the parents and children who participated in this research.

References

- Bates E. Comprehension and production in early language development. *Monographs of the Society for Research in Child Development*. 1993; 58(3-4):222-242. [PubMed: 8366873]
- Bock K. Toward a cognitive psychology of syntax: Information processing contributions to sentence processing. *Psychological Review*. 1982; 89:1-47.
- Brown, RW. *A first language; the early stages*. Cambridge, Mass: Harvard University Press; 1973.
- Bryk, AS.; Raudenbush, SW. *Hierarchical linear models: Applications and data analysis methods*. Newbury Park: Sage; 1992.
- Camarata SM, Nelson KE, Camarata MN. Comparison of conversational-recasting and imitative procedures for training grammatical structures in children with specific language impairment. *Journal of Speech and Hearing Research*. 1994; 37(6):1414-1423. [PubMed: 7877298]
- Cronbach, L.; Snow, R. *Aptitudes and instructional methods*. New York: Irvington; 1977.
- Cronbach, LN.; Gleser, GC.; Norda, H.; Rajaratnam, N. *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: Wiley & Sons; 1972.
- Ellis-Weismer S, Evans JL. The role of processing limitations in early identification of specific language impairment. *Topics in Language Disorders*. 2002; 22:15-29.
- Frangakis CE, Rubin DB. Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika*. 1999; 86:889-907.
- Goldman, R.; Fristoe, M. *Goldman Fristoe 2 Test of Articulation*. Circle Pines, MN: American Guidance Service; 2000.
- Hancock, T.; Kaiser, A. Enhanced milieu teaching. In: McCauley, R.; Fey, M., editors. *Treatment of language disorders in children*. Communication and language intervention series. Baltimore, MD: Brookes; 2006. p. 203-236.
- Kaiser, A.; Yoder, PJ. Unpublished data. Vanderbilt University; 1992. A comparison of milieu language teaching and responsive interaction as implemented by parents. Reported in Yoder, PJ. Summary of three studies that compare language interventions for preschoolers with disabilities. Paper presented at the Division of Early Childhood; Washington, D.C. 1992 December.
- Kaiser, A.; Yoder, P.; Keetz, A. Evaluating milieu teaching. In: Warren, S.; Reichle, J., editors. *Causes and effects in communication and language intervention*. Communication and language intervention series. Baltimore, MD: Brookes; 1992. p. 9-47.
- Kraemer H, Kiernan M, Essex M, Kupfer D. How and why criteria defining moderators and mediators differ between the Baron and Kenny and MacArthur approaches. *Health Psychology*. 2008; 27:101-108.
- Kraemer H, Wilson T, Fairburn C, Agras S. Mediators and moderators of treatment effects in randomized clinical trials. *Archives of General Psychiatry*. 2002; 59:877-883. [PubMed: 12365874]

- Law J, Garrett Z, Nye C. The efficacy of treatment for children with developmental speech and language delay/disorder: A meta-analysis. *Journal of Speech, Language, Hearing Research*. 2004; 47:924–943.
- Leonard, L. *Children with specific language impairment*. MIT Press; Cambridge, MA: 1998.
- Leonard L, Camarata S, Pawlowska M, Brown B, Camarata M. Acquisition of tense and agreement morphemes by children with specific language impairment during intervention: Phase 3. *Journal of Speech, Language, Hearing Research*. 2008; 51:120–125.
- Long, S.; Fey, M.; Channell, R. *Computerized Profiling (CP) Version 9.2.7 (MS-DOS)*. [Computer software]. Cleveland, OH: Department of Communication Sciences, Case Western Reserve University; 2000.
- Miller, JF. *Assessing language production in children: Experimental procedures*. Boston: Allyn and Bacon; 1991.
- Miller JF, Chapman RS. The relation between age and mean length of utterance in morphemes. *Journal of Speech and Hearing Research*. 1981; 24(2):154–161. [PubMed: 7265928]
- Miller, J.; Chapman, R. *SALT: Systematic analysis of language transcripts: Basic SALT programs*. Language Analysis Laboratory, University of Wisconsin; 1993.
- Pepper, J.; Weitzman, E. *It Takes Two to Talk: A practical guide for parents of children with language delays*. Toronto, Ontario, Canada: The Hanen Centre; 2004.
- Preacher KJ, Curran PJ, Bauer DJ. Computational tools for probing interaction effects in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics*. 2006; 31:437–448.
- Proctor-Williams K, Fey ME, Loeb DL. Parental recasts and production of copulas and articles by children with specific language impairment and typical language. *American Journal of Speech and Language Pathology*. 2001; 10:155–168.
- Roid, G.; Miller, L. *Leiter International Performance Scale - Revised*. Wood Dale, IL: Stoelting, Co; 1997.
- Scarborough H. Index of productive syntax. *Applied Psycholinguistics*. 1990; 11:1–22.
- Scarborough HS, Rescorla L, Tager-Flusberg H, Fowler A, Sudhalter V. The relation of utterance length to grammatical complexity in normal and language-disordered groups. *Applied Psycholinguistics*. 1991; 12:23–45.
- Schopler, E.; Reichler, R.; Renner, B. *Childhood Autism Rating Scale*. Los Angeles: Western Psychological Services; 1999.
- Singer, J.; Willett, J. *Applied longitudinal data analysis*. Oxford, UK: Oxford University; 2003.
- Stevens G, Cho JH. Socioeconomic indexes and the new 1980 census occupational classification scheme. *Social Science Research*. 1986; 14:142–168.
- Tapp, J. *ProCoderDV*. [computer software]. Nashville, TN: Vanderbilt Kennedy Center; 2003.
- Warren, SF. Enhancing communication and language development with milieu teaching procedures. In: Cipani, E., editor. *A guide for developing language competence in preschool children*. Springfield, IL: Charles Thomas; 1991. p. 68-93.
- Warren, S.; Rogers-Warren, A. *Teaching functional language*. Austin, TX: Pro-Ed; 1985.
- Yoder PJ, Camarata S, Gardner E. Treatment effects on speech intelligibility and length of utterance in children with specific language and intelligibility impairments. *Journal of Early Intervention*. 2005; 28:34–49.
- Yoder PJ, Compton D. Identifying predictors of treatment response. *Mental Retardation and Developmental Disabilities Research Reviews*. 2004; 10:162–168. [PubMed: 15611985]
- Yoder PJ, Kaiser AP, Goldstein H, Alpert C, Moussetis L, Kaczmarek L, Fischer R. An exploratory comparison of Milieu teaching and Responsive Interaction in the classroom. *Journal of Early Intervention*. 1995; 19:218–242.
- Yoder P, Lieberman R. What does teaching declaratives tell us about the criteria by which we can judge the developmental importance of treatment outcomes? *Evidence-Based Communication Assessment and Intervention*. 2008; 2:225–235. PMID pending. [PubMed: 21243112]

- Yoder, PJ.; McDuffie, A. Treatment of primary language disorders in early childhood: Evidence of efficacy. In: Accardo, P., editor. Disorders of language development. Baltimore: York Press; 2002. p. 151-177.
- Yoder PJ, Spruytenburg H, Edwards A, Davies B. Effect of verbal routine contexts and expansions on gains in the mean length of utterance in children with developmental delays. *Language, Speech, Hearing Services in Schools*. 1995; 26:21–32.
- Yoder, PJ.; Symons, F. *Observational measurement of behavior*. New York: Springer; 2010.
- Zimmerman, I.; Steiner, B.; Pond, R. *Preschool Language Scale-3rd Edition*. San Antonio: Psychological Corporation; 1992.

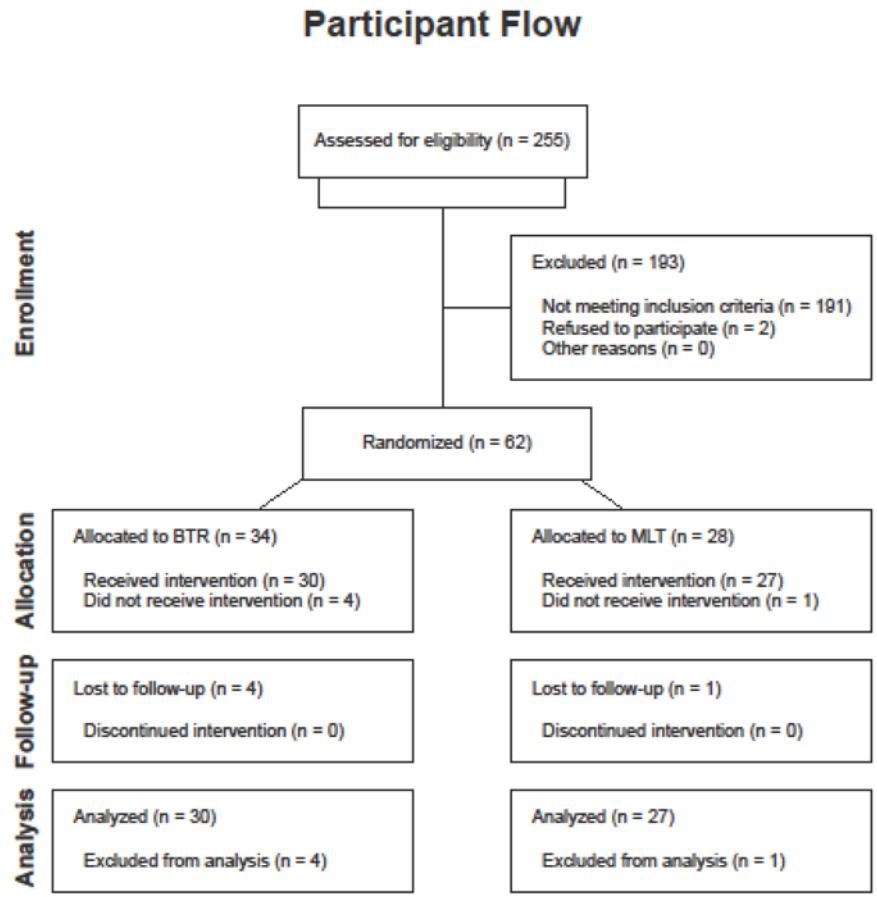


Figure 1. Participant Flow for Randomized Group Experiment

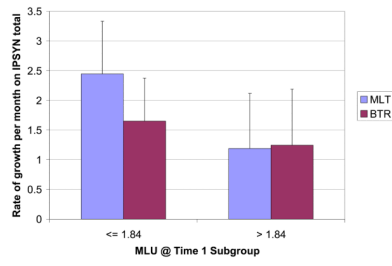


Figure 2.
Illustration of the Statistical Interaction Between MLU at Time 1 and Treatment on Grammatical Growth

Table 1

Comparability of Treatment Groups on Moderator, Time 1 Measure of IPSyn and Other Descriptive Variables.

Descriptors	MLT (27) BTR (30)		<i>p</i>	<i>d</i>
	M(SD)	M(SD)		
IPSyn total score ^a	37(13)	37(13)	.98	0.00
MLU ^a	2.0(.5)	2.0(.5)	.87	0.06
No. different word roots ^a	138(61)	137(48)	.96	0.02
Goldman-Fristoe SS ^b	91(11)	90(17)	.75	0.07
Expressive Language PR ^c	5.5(3)	5.0(3)	.75	0.09
Nonverbal Cognition SS ^d	103(13)	98(9)	.11	0.45
Age in years	3.6(.5)	3.6(.7)	.91	0.03
Parental formal education ^e	6.8(1.3)	7.1(.8)	.31	-.28
Parental occupational status ^f	53(21)	54(20)	.82	-.05

^aTwo 20-minute language samples.

^bGoldman-Fristoe Test of Test of Articulation 2 (Goldman & Fristoe, 2000).

^cPreschool Language Scale III (Zimmerman, et al., 1992).

^dLeiter International Performance Scale-Revised (Roid & Miller, 1997).

^eNine point scale of years in school with 7 = at least 4 years in college.

^fStevens and Cho, 1986 scale of occupational status with population mean of 29 (SD = 23).

SS = standard score. PR = percentile ranking.

Table 2

Means (M) and Standard Deviations (SD) for IPSyn Total Score ^a in SLI Sample(n) by Treatment Group and Assessment Period

	<u>MLT (27)</u>	<u>BTR (30)</u>
Period	M(SD)	M(SD)
1	38.4(13.3)	38.5(12.5)
2	43.7(13.7)	44.2(14.4)
3	46.6(13.5)	46.7(13.7)
4	51.2(11.2)	50.1(11.8)
5	52.6(10.9)	50.4(15.2)
6	55.9(8.3)	55.2(12.1)

^aFrom language samples.

Note: IPSyn = Index of Productive Syntax (Scarborough, 1990).

Table 3

Fixed (Top) and Random (Bottom) Effects Estimates of the Predictors of Grammatical Growth in the SLI Group

Parameter	Predictor	Coefficient(SE)	<i>p</i> value
Fixed			
Intercept (initial IPSYN total score)			
	Intercept	38.9(.83)	<.000
	Treatment	-1.36(1.7)	.42
	MLU at Time 1	21.9(1.8)	<.001
	MLU × Treatment	3.0(3.7)	.42
Slope (i.e., growth rate)			
	Intercept	1.5(.11)	<.001
	Treatment	.24(.22)	.26
	MLU at Time 1	-.86(.22)	<.001
	MLU × Treatment	-.87(.43)	.049
Random			
	Intercept	14.05	.008
	Slope	.32	.14
	Level 1	6.8	