RESEARCH ARTICLE

Behavioral Predictors of Improved Speech Output in Minimally Verbal Children With Autism

Karen Chenausky, Andrea Norton, Helen Tager-Flusberg, and Gottfried Schlaug

We investigated the relationship between eight theoretically motivated behavioral variables and a spoken-language-related outcome measure, after 25 sessions of treatment for speech production in 38 minimally verbal children with autism. After removing potential predictors that were uncorrelated with the outcome variable, two remained. We used both complete-case and multiple-imputation analyses to address missing predictor data and performed linear regressions to identify significant predictors of change in percent syllables approximately correct after treatment. Baseline phonetic inventory (the number of English phonemes repeated correctly) was the most robust predictor of improvement. In the group of 17 participants with complete data, ADOS score also significantly predicted the outcome. In contrast to some earlier studies, nonverbal IQ, baseline levels of expressive language, and younger age did not significantly predict improvement. The present results are not only consistent with previous studies showing that verbal imitation and autism severity significantly predict spoken language outcomes in preschool-aged minimally verbal children with autism, but also extend these findings to older minimally verbal children with autism. ***Autism Research*** *2018, 11: 1356–1365.* © 2018 International Society for Autism Research, Wiley Periodicals, Inc.

Lay Summary: We wished to understand what baseline factors predicted whether minimally verbal children with autism would improve after treatment for spoken language. The outcome measure was change in percentage (%) syllables approximately correct on a set of 30 two-syllable words or phrases. Fifteen were both practiced in treatment and tested; the remainder were not practiced in treatment, but only tested, to assess how well children were able to generalize their new skills to an untrained set of words. Potential predictors tested were sex, age, expressive language, phonetic inventory (the number of English speech sounds repeated correctly), autism severity, and nonverbal IQ. Phonetic inventory and (for some children) autism severity predicted children's posttreatment improvement. Nonverbal IQ and expressive language ability did not predict improvement, nor did younger age, suggesting that some older children with autism may be candidates for speech therapy.

Keywords: minimally verbal; autism spectrum disorder; spoken language; intonation; longitudinal data analysis

Introduction

In recent years, an increasing number of studies have been devoted to the minimally verbal (MV) segment of the autism spectrum disorder (ASD) population–children who fail to acquire phrase speech by age 5 [Tager-Flusberg & Kasari, 2013]. Given that upward of 25% of children with ASD fall into this category [Mawhood, Howlin, & Rutter, 2000; Anderson et al., 2007; Norrelgen et al., 2015; Rose, Trembath, Keen, & Paynter, 2016], of particular importance are investigations that target spoken language acquisition in this population. Such research has shown that better expressive language skills are associated both with better long-term outcomes [Howlin, Mawhood, & Rutter, 2000] and fewer maladaptive behaviors [Baghdadli, Pascal, Grisi, & Aussilloux, 2003; Dominick, Davis, Lainhart, Tager-Flusberg, & Folstein, 2007; Hartley, Sikora, & McCoy, 2008; Matson & Rivet, 2008]. Because maladaptive behaviors decrease and long-term outcomes improve when children learn to communicate more successfully [Buschbacher & Fox, 2003], efforts to identify and understand the factors that predict improvement in spoken language have become increasingly important as we seek to develop more effective treatments for these MV children.

Auditory-Motor Mapping Training

The current analysis was performed in the context of a completed proof-of-concept study and an ongoing

Form the Music, Neuroimaging, and Stroke Recovery Laboratory, Beth Israel Deaconess Medical Center, Boston, MA (K.C., A.N., G.S.); Department of Neurology, Harvard Medical School, Boston, MA (K.C., G.S.); Center for Autism Research Excellence, Department of Psychological and Brain Sciences, Boston University, Boston, MA (K.C., H.T-F.)

Received September 13, 2017; accepted for publication July 1, 2018
Address for correspondence and reprints: Gottfried Schlaug, Department of Neurology, Harvard Medical School, 25 Shattuck Street, Boston, MA 02115. E-mail: gschlaug@bidmc.harvard.edu
Published online 19 September 2018 in Wiley Online Library (wileyonlinelibrary.com)
DOI: 10.1002/aur.2006

randomized clinical trial, comparing auditory-motor mapping training (AMMT), a novel treatment that uses intonation (singing) and rhythmic hand tapping to facilitate sound-motor mapping and/or to improve speech output in MV children with ASD, with speech repetition therapy (SRT), a control treatment that involves neither intonation nor hand tapping. AMMT is one of a small number of music-based treatments that have recently begun to be used effectively for teaching language and social skills to children with ASD [see e.g.,Lim, 2010 ; Lim & Draper, 2011 ; Paul et al., 2015]. In an earlier proof-of-concept study involving AMMT alone, significant improvement in production of two-syllable words and phrases was observed over 40 treatment sessions in six MV children with ASD [Wan et al., 2011], with levels of improvement ranging from 8% to 71% across participants. More recently, Chenausky, Norton, Tager-Flusberg, and Schlaug [2016] compared the effectiveness of AMMT and SRT in a group of 30 MV children with ASD [including the six from Wan et al., 2011]. Ten of the 30 subjects received 40 AMMT sessions, 13 received 25 AMMT sessions, and 7 received 25 sessions of SRT. For the current report, the assessment after 25 sessions was used as a common basis of comparison across these three samples, which were assembled during three different phases of this pilot research. Compared to baseline, AMMT participants improved by an average of +19.4% syllables approximately correct (range 26.7% to +48.4%), while SRT participants improved by just +3.6% on the same measure (range 8.4% to +13.4%). In addition, significantly more AMMT participants than SRT participants (83% vs. 14%) were "responders"—they showed a statistically significant improvement after 25 therapy sessions. In a subsequent study, Chenausky et al. [2017] showed that, in a comparison between matched pairs of more-verbal and MV children with ASD, the AMMT-treated child in each pair showed greater improvement in speech output than the SRT-treated children, with greater effect sizes for the more-verbal child than the MV child after the same number of sessions.

We sought to investigate factors that may have contributed to the improvement that children experienced, regardless of treatment group. In Chenausky et al. [2016] we performed a preliminary correlational analysis to investigate the relationship of three baseline variables to change scores after 25 therapy sessions in the AMMT participants. Here, we expand on that work by including a larger number of AMMT and SRT participants, considering a larger number of potential predictors, and using regression analyses to quantify the effect of those predictors on the outcome measure "change in percentage (%) syllables approximately correct".

Selecting Potential Predictors of Improvement in Speech Output

Based on the conclusions of earlier research, we chose several theoretically-motivated predictor variables measured at baseline that have all been shown to be significantly related to expressive language outcomes: Sex [Carter, Black, Tewani, Connolly, & Tager-Flusberg, 2007; Reinhardt, Wetherby, Schatschneider, & Lord, 2015]; chronological age [Fenske, Zalenski, Krantz, & McClannahan, 1985; Venter, Lord, & Schopler, 1992; but see Smith, Mirenda, & Zaidman-Zait, 2007]; measures of expressive language (e.g., expressive vocabulary, intelligible words or word approximations) [Smith et al., 2007; Thurm, Lord, Lee, & Newschaffer, 2007; Ellis Weismer & Kover, 2015], phonetic inventory [Thurm et al., 2007; Smith et al., 2007; Wetherby, Watt, Morgan, & Shumway, 2007; Yoder, Watson, & Lambert, 2015], autism severity [Ellis Weismer & Kover, 2015], and nonverbal IQ [Venter et al., 1992; Thurm et al., 2007; Ellis Weismer & Kover, 2015].

Our goal was to determine which of these variables predicted the magnitude of improvement in speech output in a group of MV children with ASD after 25 sessions of treatment.

Methods

Participants

Participants included 38 MV children with ASD between the ages of 3;5 and 10;8 participating in two IRB-approved studies, the first a proof-of-concept study providing pilot data for the second, an ongoing randomized controlled clinical trial (RCT) comparing the efficacy of AMMT and SRT. Of the 38 participants from the combined investigations (pilot, proof-of-concept, and RCT), one child (male) received 60 sessions of AMMT, 10 children (three female) received 40 sessions of AMMT, 16 children (one female) received 25 sessions of AMMT, and the remaining 11 (three female) received 25 sessions of SRT. All children were assessed at least 3 times at baseline and again after the 10th, 15th, 20th, and 25th treatment sessions. For comparison across children, the post-25 (P25) assessment session was used, as the common element in the various phases of this research is that all children received at least 25 treatment sessions and had an assessment after the 25th treatment session. Each child's best baseline score was compared with their P25 score. Children were recruited from multiple autism clinics and resource centers serving the Greater Boston area, and via autism networks online. Both protocols were approved by the Institutional Review Board of Beth Israel Deaconess Medical Center, and parents of all participants gave written informed consent prior to enrollment.

Diagnostic status was confirmed by a Childhood Autism Rating Scale [CARS; Schopler, Reichler, & Rochen-Renner, 1988] score greater than 30 or an Autism Diagnostic Observation Schedule [ADOS; Lord, Rutter, DiLavore, & Risi, 2002] score greater than 12. MV status, confirmed by parent report and child performance during initial assessments, was defined as using fewer than 20 intelligible words and no productive syntax [Tager-Flusberg & Kasari, 2013]. To be included in the studies, children had to be able to correctly repeat at least two speech sounds, participate in table-top activities for at least 15 min at a time, follow one-step commands, and imitate simple gross and oral motor movements, such as clapping hands and opening mouth. Children continued with their regular school programs during the study but did not participate in any speech therapy activities or new treatments outside of school. Aside from ASD, participants had no other major neurological conditions (e.g., tuberous sclerosis), motor disabilities (e.g., cerebral palsy), sensory disabilities (e.g., hearing or sight impairment), or genetic disorders (e.g., down syndrome) that might explain their MV state.

Interventions

The theoretical basis and structure of AMMT is described in detail in Wan and Schlaug [2010], Wan et al. [2011], Chenausky et al. [2016], and Chenausky, Norton, and Schlaug [2017]. Here, we briefly outline the major characteristics of AMMT and SRT. Both treatments used the same set of 30 bisyllabic words or phrases referring to people, objects, or activities familiar to children (e.g., "mommy," "cookie," "bye-bye"). Children's performance producing these stimuli was assessed over multiple (3–5) baselines and again after 10, 15, 20, and 25 therapy sessions. During therapy sessions, only 15 of these stimuli were practiced; these comprised the trained set of stimuli. The remaining 15 stimuli (untrained) were only presented during assessment sessions; their purpose was to assess the extent to which children were able to generalize the skills they learned in treatment to a set of words they had not practiced.

Treatment sessions lasted approximately 45 min and took place 5 days/week for 25 sessions. During therapy, children had multiple opportunities to produce each stimulus and to receive corrective feedback on their performance across five treatment steps. These steps range from full therapist support (child and therapist produce the target stimulus in unison) to independent production (child produces the target alone after a cue). During assessments, the same steps and prompts were used, but no corrective feedback was given.

AMMT and SRT differ in that AMMT involves tapping tuned drums while simultaneously intoning, or singing, the stimuli. As discussed in the references cited above, the combination of intonation and bimanual movement is thought to facilitate the acquisition of speech by engaging an auditory-motor feedforward-feedback network and by facilitating the mapping of sounds to articulatory actions. Among the other proposed mechanisms, the intonation component also increases participants' interest and attention during the intervention. SRT, by contrast, does not involve drums, bimanual movement, or intonation. In this sense, SRT is a treatment-as-usual comparison that keeps the dose (length and number of sessions, number and type of stimuli, and opportunities to produce each one) constant between the two conditions.

Measures

Outcome measure. The outcome measure used in this study is *change in percentage (%) syllables approximately correct*. The use of a perceptual-based measure of word production is implicit in previous treatment literature [e.g., Rogers et al., 2006; Yoder & Stone, 2006; Paul, Campbell, Gilbert, & Tsiouri, 2013] and is clinically meaningful as a proxy for the degree to which a child's communication partner is able to identify or understand the word that is produced. Also implicit in the previous literature is that a child's production of a word need not be a perfect imitation of the adult target to be understood. For example, Yoder and Stone [2006] defined their outcome measure, a spoken communication act, as "an utterance that contains one or more intelligible word approximation(s)." (p. 704).

In the case of percentage (%) syllables approximately correct, we used an explicit rubric for determining whether a child's production was a sufficiently accurate approximation of the adult target. All of a child's responses during Baseline and probe assessments were phonetically transcribed by raters blind to the study time point. Each syllable in the 30 bisyllabic words/phrases was scored as "approximately correct" or not based on the number of phonetic features the child's phonemes shared with the adult target. To be approximately correct, (1) the initial consonant of the syllable must have shared at least two of three features (place, manner, voicing) with the target consonant and (2) the vowel of the syllable must have shared two features (height, backness) with the target vowel. The change score was then calculated by subtracting a child's best baseline score from their score at the post-25 assessment. Best Baseline was defined as the baseline session during which a child produced the highest number of syllables approximately correct over the 30 bisyllabic words/phrases.

To assess inter-rater reliability, 10% of probes across participants were transcribed and scored by two independent transcribers. As reported in Chenausky et al. [2016], results yielded 68.0% agreement on *syllables approximately correct* (Cohen's $\kappa = 0.497$, $P < 0.0005$), 70.1% agreement on *consonants correct* (κ = 0.547, $P < 0.0005$), and 54.7% agreement on *vowels correct* (κ = 0.270, $P < 0.0005$). These figures are comparable to previously published figures of 77% agreement on consonants correct and 45% agreement on vowels correct for transcriptional analyses of infant babbles [Davis & MacNeilage, 1995].

Predictors. Based on previous work and the existing literature, we identified several potential predictors to include in our initial correlational analysis. These were: sex, chronological age, expressive language, phonetic inventory, autism severity, and NVIQ. All measures were collected at baseline. The measure of *expressive language* was baseline score on the Mullen Scales of Early Learning [MSEL; Mullen, 1995], Expressive Language subtest. Raw scores on the Mullen were used because all participants scored below the 1st percentile for their ages, so raw scores are more informative and yield a greater range of values than *T* scores. *Phonetic inventory* was assessed by an imitation task in which children were asked to repeat all of the consonants and vowels in English (31 total); number of correctly repeated phonemes was then used as a predictor in subsequent analyses. For *autism severity*, we used total ADOS score. Note that a severity calibration metric for total ADOS scores is available [Gotham, Pickles, & Lord, 2009] to compare autism severity across variations in IQ and language level. However, since our inclusion criteria resulted in a relatively uniform sample (all MV children, assessed with Module 1), we used total ADOS score instead of the calibrated severity score, as the former yielded a greater range of scores than the latter. Finally, for *NVIQ*, we used the visual reception score on the MSEL. Again, raw scores are reported. Table 1 details the baseline and outcome scores for all participants.

Analytic Strategy

Overall strategy. First, we performed a repeated measures ANOVA to establish that, as in previous work, there was a significant between-group treatment effect in favor of AMMT. Following this, our general analytic strategy was to construct, test, and compare a series of linear regression models relating our putative predictors to our outcome variable. However, there were two challenges to this approach: (1) there was a relatively small number of participants, and (2) some predictor data were missing because children in the earliest studies had not all been assessed with the same test instruments. This necessitated the adoption of two strategies to ensure that our conclusions from these data were valid. Stata v.14 was used for all analyses [StataCorp, 2015].

Variable selection. Our goal was to develop a regression model that quantified the relationship of our putative predictors to the outcome measure. However, a regression including 6 predictors and only 38 subjects runs the risk of overfitting (i.e., creating a model whose significant predictors are predictive of the outcome in the study sample, but not in the overall population). While overfitting does not bias estimates of the regression coefficients, it can result in models whose estimates of regression coefficient magnitude, variability, and significance are very sensitive to small, meaningless fluctuations in data values. Kleinbaum, Kupper, Nizam, and Rosenberg [2014] suggest 5–10 observations per predictor as a rule of thumb. For a sample size of 38, this would mean 4–8 predictors in a model. Thus, we first ran initial regression models including all six predictors. Predictors that were not significant were removed, and second set of

Table 1. Subject characteristics

	AMMT[e]	SRT[f]
Sex	4 F, 23 M	3 F, 8 M
Age (yr; mo)		
μ	6;8	6;2
±SD	±1;10	±1;6
n	27	11
Phonetic Inventory[a]		
μ	7.4	8.7
±SD	±4.7	±6.7
n	27	11
ADOS[b]		
μ	19.5	21.6
±SD	±3.2	±3.4
n	15	9
MSEL EL[c]		
μ	10.8	11.7
±SD	±1.9	±3.9
n	13	7
MSEL VR[d]		
μ	29.1	31.7
±SD	±8.6	±11.3
n	16	10
Change in % syllables approximately correct		
μ	17.8	0.5
±SD	±18.8	±12.0
n	27	11

[a]Phonetic inventory: The number of English phonemes a child correctly repeated at baseline (max = 31).

[b]ADOS: Autism diagnostic observation schedule. Cutoff for a diagnosis of autism = 12; for autism spectrum disorder = 8.

[c]MSEL EL: Mullen scales of early learning, Expressive language subscale. Raw score reported (max = 50).

[d]MSEL VR: Mullen scales of early learning, Visual reception subscale. Raw score reported (max = 50).

[e]AMMT: Auditory-motor mapping training.

[f]SRT: Speech repetition therapy.

regression models was run that included just the predictors that were significant, plus interaction terms to assess whether the association of those predictors varied as a function of treatment group. This procedure allowed us to construct more parsimonious regression models [Kleinbaum et al., 2014].

Dealing with missing data. The strategy described above, however, is complicated by the fact that we lacked ADOS and MSEL scores for the very first participants in the study. These data are *missing completely at random* [MCAR; Chen, Ibrahim, Chen, & Senchaudhuri, 2008, Graham, 2009] as which values are missing depends on neither the predictor nor the outcome variables. Thus, parameters such as the mean and variance of the overall sample can be estimated from the complete cases, and regression parameter estimates will be unbiased (i.e., close to the actual population values).

Complete case analysis versus multiple imputation. Two methods of dealing with missing data were used: complete case analysis and multiple imputation. *Complete case analysis* means analyzing only cases for which all data points are available. This method reduces the sample size and overall statistical power and, thus, the number of potential predictors it is possible to test. However, as long as the data are MCAR, complete case analysis does not necessarily result in biased parameter estimates. *Multiple imputation* is the process of generating plausible values for the missing data points multiple times and then aggregating the results, taking advantage of known characteristics of the existing data, such as mean and variance. As the imputed variables in this case are functions only of baseline covariates, multiple imputation introduces no bias into the regression parameter estimates [White & Thompson, 2005]. Both the complete-case and multiple-imputation analyses are reported and interpreted here, as both provide useful information about the sample and population under discussion.

Results

Establishing a Treatment Effect

A repeated-measures ANOVA was performed on percentage (%) syllables approximately correct, with time (baseline vs. P25) and stimuli (trained vs. untrained) as within-subjects factors and treatment (AMMT vs. SRT) as between-subjects factors.

There was a significant main effect of time, $F(1,36) = 8.950$, $P = 0.005$, indicating that, on average, the participants in this study improved between baseline and P25. Mean baseline score was 31.0% syllables approximately correct [standard error(SE) 3.3], compared to the mean score at P25, which was 40.2% (SE 4.7). There was also a significant main effect of stimuli, $F(1,36) = 30.323$, $P < 0.0005$. Mean percentage (%) syllables approximately correct for trained stimuli was 39.2% (SE 3.8), compared to 32.1% for untrained stimuli (SE 4.7). There was no significant main effect of Treatment, indicating that the two groups did not show a consistent difference across timepoints.

There was, however, a significant time × treatment interaction, $F(1,36) = 7.924$, $P = 0.008$. For the AMMT group, mean baseline score was 24.5% (SE 3.5) and mean P25 score was 42.3% (SE 5.0), Cohen's $d = 0.81$ (large). For the SRT group, mean baseline score was 37.6% (SE 5.6) and mean P25 score was 38.1% (SE 7.8), $d = 0.02$ (negligible). Thus, AMMT participants improved significantly more than SRT participants. There were no other significant two-way effects.

Finally, there was a significant time × treatment × stimuli interaction, $F(1,36) = 8.095$, $P = 0.007$. AMMT participants improved by a mean of 19.9% points on trained stimuli ($d = 0.86$, large) and a mean of 15.5 on untrained stimuli ($d = 0.71$, medium). SRT participants decreased by a mean of 3.6% points on trained stimuli ($d = 0.16$, small) and improved by a mean of 4.7 on untrained stimuli ($d = 0.22$, small).

A repeated-measures ANOVA was also performed on percentage (%) syllables approximately correct for the trained and untrained stimuli separately, again with time (baseline vs. P25) as a within-subjects factor and treatment (AMMT vs. SRT) as a between-subjects factor. Means, standard errors of mean (*SEM*), and 95% confidence intervals (CI) for baseline and P25 scores on trained and untrained stimuli for both groups appear in Table 2. For trained stimuli, there was a significant main effect of time ($F(1,36) = 5.310$, $P = 0.027$) and a significant time × treatment interaction ($F(1,36) = 11.107$, $P = 0.002$). For untrained stimuli, there was also a significant main effect of time ($F(1,36) = 11.890$, $P = 0.001$), but

Table 2. Percentage syllables approximately correct, by treatment group, stimulus type, and timepoint

Group	Stimuli	Timepoint	Mean (SE[d])	95% CI[e]
AMMT[a] (n = 27)	Trained	Baseline	26.5 (3.7)	18.9–34.2
		P25[c]	46.4 (5.2)	36.0–57.0
	Untrained	Baseline	22.5 (3.5)	15.3–29.6
		P25	38.1 (5.0)	27.9–48.4
SRT[b] (n = 11)	Trained	Baseline	43.6 (6.0)	31.7–55.6
		P25	40.0 (8.1)	23.6–56.4
	Untrained	Baseline	31.5 (5.5)	20.3–42.7
		P25	36.2 (7.9)	20.0–52.2

[a]AMMT: Auditory-motor mapping training.
[b]SRT: Speech repetition therapy.
[c]P25: Post 25 sessions assessment.
[d]*SE*: Standard error of the mean.
[e]CI: Confidence interval.

the time × treatment interaction was nonsignificant (F (1,36 = 3.443, P = 0.072). These results indicate that there is an effect of treatment group on trained but not untrained stimuli.

Complete Case Analyses

The next step in our analysis was to fit a linear regression model predicting change in percentage (%) syllables approximately correct with all six potential predictor variables, including only participants whose datasets were complete: a set of 12 AMMT participants and 5 SRT participants. To establish that there were no differences at baseline between the complete-case group and the group with incomplete baseline data on the other measures, we performed a series of two-tailed t tests with $\alpha = 0.05$ on sex, chronological age, and baseline score of percentage (%) syllables approximately correct. These were uncorrected for multiple comparisons, as we wished to identify any baseline differences that might be present. All P values were greater than 0.1, demonstrating that there were no significant differences between complete and incomplete cases on any of the baseline measures for which they all had data.

The overall regression model was significant, $F(6,10) = 5.97$, $P < 0.007$, $R^2 = 0.782$, adjusted $R^2 = 0.651$. However, only ADOS score, chronological age, sex, and phonetic inventory significantly predicted change in percentage (%) syllables approximately correct and were retained in the next step.

Next, a regression model including ADOS score, chronological age, phonetic inventory, sex, and interaction terms between these variables and treatment was fit. Again, the overall model was significant, $F(9,14) = 4.14$, $P = 0.009$, $R^2 = 0.727$, adjusted $R^2 = 0.552$. In this case, only ADOS score and phonetic inventory were significant; no interaction terms were significant. Regression parameter estimates and standard error for both analyses are provided in Table 2.

Multiple Imputation Analyses

Next, we describe the results from the multiple imputation analyses (Table 4). The 20 imputations were used for the missing ADOS, MSEL EL, and MSEL VR scores; data from the imputations was aggregated and used in the regression. A multivariate normal (mvn) distribution method was used. In addition, a correlation analysis was performed to determine whether auxiliary variables (variables in the data set that are either correlated with the missing variables or believed to be associated with missingness) should be included in the analysis. As mentioned earlier, no variables were found to be associated with missingness. The correlation analysis showed that no baseline variables were significantly correlated with any others (all $P > 0.05$). Therefore, no auxiliary variables were included in the multiple imputation analysis.

As before, the initial regression model included all six predictors. The overall model was not significant, $F(6,26.8) = 1.29$, $P = 0.294$. Only phonetic inventory significantly predicted change in percentage (%) syllables approximately correct. Therefore, the second model included phonetic inventory and a treatment × phonetic inventory interaction term. This model was significant, $F(3,32.2) = 6.66$, $P = 0.001$, $R^2 = 0.370$, adjusted $R^2 = 0.315$.

Regression parameter estimates and standard errors for the multiple imputation analyses are reported in Table 3.

Discussion

In this study, we examined potential predictors of improvement in a measure of spoken language, Change in percentage (%) syllables approximately correct, in a group of 38 school-aged MV children with ASD. Several findings emerged from the analysis.

First, a repeated-measures ANOVA comparing percentage (%) syllables approximately correct at baseline and after 25 therapy sessions showed that there was a significant time × treatment effect. This indicates that AMMT is responsible for at least some of the improvement in our

Table 3. Regression model: complete case analysis. Top: all predictors. Bottom: significant predictors plus interaction terms

	β[c]	SE[d] (β)	P value
AMMT + SRT (n = 24)			
All predictors			
ADOS score	−3.904	1.109	0.006
CA[a]	5.370	1.611	0.008
EL	−2.062	1.698	–
NVIQ[b]	−0.236	0.327	–
Phonetic inventory	1.852	0.550	0.007
Sex	−40.252	11.631	0.006
Constant	71.909	27.025	0.024
Significant predictors plus interaction terms			
ADOS score	−2.809	1.046	0.018
CA	1.730	1.564	–
Phonetic inventory	1.601	0.693	0.037
Sex	4.049	13.102	–
Treatment × ADOS	1.662	1.692	–
Treatment × CA	−2.802	3.176	–
Treatment × Phonetic inventory	1.253	0.954	–
Treatment × Sex	−1.036	0.954	–
Constant	51.805	22.082	0.034

[a]CA: Chronological age.
[b]NVIQ: Nonverbal IQ (i.e., Mullen scales of early learning visual reception raw score).
[c]β: Regression coefficient (unstandardized).
[d]*SE*: Standard error.

Table 4. Regression model: multiple imputation analysis. Top: all predictors. Bottom: significant predictors plus interaction terms

	β	*SE*(β)	*P* value
	AMMT + SRT (n = 38)		
	All predictors		
ADOS Score	−1.443	1.182	–
CA	2.484	1.790	–
EL	−0.145	1.647	–
NVIQ	−0.359	0.406	–
Phonetic inventory	1.335	0.610	0.038
Sex	−6.613	8.467	–
Constant	31.624	30.509	–
	Significant predictors plus interaction terms		
Phonetic inventory	2.054	0.652	0.004
Treatment × Phonetic inventory	−1.663	0.980	–
Constant	2.562	5.695	–

For abbreviations, see captions for previous tables.

participants' speech production and strongly suggests that it can bring about changes in the consonant inventories of minimally verbal children with ASD that generalize to untrained words and phrases. Research focused on changes in consonant inventory has been identified by Yoder et al. [2015] as an area of high importance. How best to foster the use of newly developed speech production skills in spontaneous communication for minimally verbal children with ASD is a separate issue to be explored in future studies.

Next, most of our theoretically motivated predictors did not significantly predict our outcome variable. The only variable to consistently emerge as a significant predictor of change in percentage (%) syllables approximately correct was phonetic inventory at baseline. In the complete case analysis, which included the 24 participants with complete data, chronological age, and ADOS score were additional significant predictors of our outcome variable. Sex and baseline measures of expressive language and NVIQ did not significantly predict change in percentage (%) syllables approximately correct. In addition, the use of both complete-case and multiple imputation analyses provided information about the relationship of phonetic inventory, chronological age, and ADOS score that one analysis alone did not respond. By using both of these analyses, we were able to gain a more complete picture of how these two predictors were related to our outcome variable. Significant values of the regression parameter for phonetic inventory ranged from 1.3 to 2.1, meaning that for every extra phoneme a child could repeat correctly at baseline, we could expect a 1.3%–2% increase in the amount of improvement he or she showed after 25 sessions of therapy. The significant value of the regression parameter for ADOS score was approximately −4, meaning that a one-unit increase in ADOS score was associated with a 4% decrease in the amount of improvement a child showed after treatment.

It is also instructive to examine the adjusted R^2 values from the two more parsimonious regression analyses, as these indicate how much of the variability in the outcome measure is because of variation in the (significant) predictors. For the parsimonious complete-case analysis, phonetic inventory and ADOS score together accounted for 73% of the variance in change in percentage (%) syllables approximately correct, for a sample of 17 participants with complete data. In the parsimonious multiple-imputation analysis (38 participants), phonetic inventory accounted for 32% of the variance in the outcome measure. Compare these figures to those of Paul et al. [2013], whose regression models predicting expressive language accounted for between 30% and 47% of the variance in outcome for a group of 22 participants; and with those from Yoder et al. [2015], whose models accounted for approximately 52% of the variance in outcome for 87 participants. That is, in each case, between one-third and three-quarters of the variance in outcome was accounted for by the significant predictors. Our results must be interpreted in the context of our relatively small sample, which does not allow us to answer the question of what is responsible for the remainder of the variance, but this is an appropriate focus for future studies with more participants.

Finally, several potential predictors were not found to be significant in predicting change in percentage (%) syllables approximately correct after 25 treatment sessions in our participants. Specifically, Baseline scores of EL and NVIQ were not significant predictors. Sex was only significant in the unparsimonious complete-case analysis. Sex was coded 0 for female and 1 for male, and the negative regression parameter for sex in this model meant that being female was associated with greater improvement than being male in this small group of participants. The fact that neither EL nor NVIQ significantly predicted outcome may suggest something about the mechanisms responsible for our participants' minimally verbal status: being minimally verbal may not be due solely to general intellectual impairment or expressive *language* impairment. Yet unknown factor may also play a role and, thus, should be the target of future research.

The finding that speech delay, NVIQ, and language impairment may be separable and independent is not new; Rice [2016] discusses the idea that accounting for the full range of developmental communication outcomes must at least consider language skill and nonverbal intelligence as independent factors. The data shown was discusses that, in a demographically diverse sample of American kindergartners, 5% experienced both low cognition (IQ < 85) and impaired language (standard score < 80). Yet the prevalence of speech delay in this group was only 0.77%—far from universal. Thus, especially given the small size of this study and the associated risk of Type 2 errors (missing an effect that is significant in the population), further research should investigate the extent to which language impairment, NVIQ, and factors specifically affecting speech production all interact to produce the minimally verbal phenotype.

Clinical Implications

The results of this study have important implications for clinical practice. First, they suggest that the ability to correctly imitate native-language phonemes may be an important factor associated with improvement in speech production for MV children with ASD. They extend previous similar findings in preschool children with ASD [Thurm et al., 2007; Smith et al., 2007; Wetherby et al., 2007; Yoder et al., 2015] to older MV children with ASD. It was not within the scope of this study to investigate the extent to which our participants were able to generalize their newfound speech production skills to spontaneous communication. However, to whatever extent that speech production underlies or contributes to expressive language, we have shown that it is modifiable through treatment. For minimally verbal children with very small phonetic inventories, then, the initial stages of therapy might include practice imitating speech sounds, in particular, intoned speech sounds.

It is also a relatively optimistic finding that chronological age was either not related to our speech outcomes, or was negatively related to it in our participants. Others [e.g., Pickett, Pullara, O'Grady, & Gordon, 2009] have noted that older minimally verbal children with ASD have been reported to attain useful speech past age 5, and in fact that the phrase or sentence level of language was achieved by some children who acquired speech as late as 11 years of age.

There is considerable appropriate concern about early identification and treatment for ASD, but the present results suggest that there may still be the possibility of improvement even for MV children with ASD who have not begun to speak by age 5. As the age was not predictive of response to treatment, we can only speculate due to the small size of our study, which may make it more likely to miss effects that are significant in the overall population. It may be that while younger children possess more latent ability to learn speech, this is balanced out by potential gains in joint attention and the ability to tolerate didactic activities for extended periods of time in older children. Regardless, older minimally verbal children with ASD should not be excluded from participation in speech interventions and, further, the potential for improving speech in older minimally verbal individuals with ASD should be the subject of future research.

Limitations and Future Work

As with most research on a population as heterogeneous as MV children with ASD, a limitation of this study is its small number of participants. MV children are particularly challenging to work with, which is why it is only fairly recently that researchers have begun to include them in studies [Wan et al., 2011; Tager-Flusberg et al., 2016; Plesa-Skwerer, Jordan, Brukilacchio, & Tager-Flusberg, 2016; Chenausky et al. 2017; Chenausky, Kernbach, Norton, & Schlaug, 2017]. In addition, the heterogeneity in this population is quite wide; approximately half of the children we screen do not meet inclusion criteria for our studies. Thus, these results need to be replicated in larger groups and in children who are receiving different forms of therapy. Given the great need for these children to acquire even a few words and that, no therapy works equally well for all children, understanding the individual characteristics that make a specific therapy appropriate for a particular MV child with ASD is an important aspect of research in this population. Future research investigating the roles of imitative ability and phonetic inventory in treatment response may deepen our understanding of their potential benefit in early intervention for autistic children at risk for being MV. Another avenue for future work already underway in our lab is the identification of comorbid conditions (e.g., motor speech disorders, such as childhood apraxia of speech) that may limit these children's ability to acquire spoken language.

Acknowledgments

We are deeply grateful for the effort of the dedicated families who committed significant amounts of their time to making this work possible. We also thank our present and former research assistants and summer students for their hard work in the collection of these data. None of the authors have relevant financial or nonfinancial relationship(s) with any products or services described, reviewed, evaluated, or compared in this report. Nancy Lurie Marks Family Foundation, Autism Speaks (7504), the National Institutes of Health (P50 DC 13027), and the American Speech-Language-Hearing Foundation.

References

Anderson, D., Lord, C., Risi, S., DiLavore, P., Shulman, C., Thurm, A., … Pickles, A. (2007). Patterns of growth in verbal abilities among children with autism spectrum disorder. Journal of Consulting and Clinical Psychology, 75(4), 594–604.

Baghdadli, A., Pascal, C., Grisi, S., & Aussilloux, C. (2003). Risk factors for self-injurious behaviors among 222 young children with autistic disorders. Journal of Intellectual Disability Research, 47(8), 622–627.

Buschbacher, P., & Fox, L. (2003). Understanding and intervening with the challenging behavior of young children with autism spectrum disorder. Language, Speech, and Hearing Services in Schools, 34, 217–227.

Carter, A., Black, D., Tewani, S., Connolly, K., & Tager-Flusberg, H. (2007). Sex differences in toddlers with autism spectrum disorders. Journal of Autism and Developmental Disorders, 27, 86–97. https://doi.org/10.1007/s10803-006-0331-7.

Chen, Q., Ibrahim, J., Chen, M., & Senchaudhuri, P. (2008). Theory and inference for regression modesl with missing responses and covariates. Journal of Multivariate Analysis, 99, 1302–1331.

Chenausky, K., Norton, A., Tager-Flusberg, H., & Schlaug, G. (2016). Auditory Motor Mapping Training: Comparing the effects of a novel speech treatment to a control treatment for minimally verbal children with autism. PLoS ONE, 11(11), e0164930. https://doi.org/10.1371/journal.pone.0164930.

Chenausky, K., Norton, A., & Schlaug, G. (2017). Auditory-motor mapping training in a more verbal child with autism. Frontiers in Human Neuroscience, 11, 426. https://doi.org/10.3389/fnhum.2017.00426.

Chenausky, K., Kernbach, J., Norton, A., Schlaug, G. (2017) White matter integrity and treatment-based change in speech performance in minimally verbal children with autism spectrum disorder. Frontiers in Human Neuroscience 11,175. doi: https://doi.org/10.3389/fnhum.2017.00175.

Davis B, MacNeilage P. (1995) The articulatory basis of babbling. Journal of Speech and Hearing Research 38, 1199–1211. PMID: 8747814.

Dominick, K., Davis, N., Lainhart, J., Tager-Flusberg, H., & Folstein, S. (2007). Atypical behaviors in children with autism and children with a history of language impairment. Research in Developmental Disabilities, 28, 145–162.

Ellis Weismer, S., & Kover, S. (2015). Preschool language variation, growth, and predictors in children on the autism spectrum. Journal of Child Psychology and Psychiatry, 56(12), 1327–1337.

Fenske, E., Zalenski, S., Krantz, P., & McClannahan, L. (1985). Age at intervention and treatment outcome for autistic children in a comprehensive intervention program. Analysis and Intervention in Developmental Disabilities, 5, 49–58.

Gotham, K., Pickles, A., & Lord, C. (2009). Standardizing ADOS scores for a measure of severity in autism spectrum disorders. Journal of Autism and Developmental Disorders, 39, 693–705.

Graham, J. (2009). Missing data analysis: Making it work in the real world. Annual Reviews of Psychology, 60, 549–576.

Hartley, S., Sikora, D., & McCoy, R. (2008). Prevalence and risk factors of maladaptive behaviour in young children with autistic disorder. Journal of Intellectual Disability Research, 52(10), 819–829.

Howlin, P., Mawhood, L., & Rutter, M. (2000). Autism and developmental receptive language disorder – a comparative follow-up in early adult life. II: Social, behavioural, and psychiatric outcomes. Journal of Child Psychology and Psychiatry, 41(5), 561–578.

Kleinbaum, D., Kupper, L., Nizam, A., & Rosenberg, E. (2014). Applied Regression Analysis and Other Multivariable Methods. Boston, MA: Cengage Learning.

Lim, H. (2010). Use of music in the applied behavior analysis verbal behavior approach for children with autism spectrum disorders. Music Therapy Perspectives, 28(2), 98–105.

Lim, H., & Draper, E. (2011). The effects of music therapy incorporated with applied behavior analysis verbal behavior approach for children with autism spectrum disorders. Journal of Music Therapy, 48(4), 532–550.

Lord, C., Rutter, M., DiLavore, P., & Risi, S. (2002). Autism diagnostic observation schedule. Los Angeles, CA: Western Psychological Services.

Matson, J., & Rivet, T. (2008). The effects of severity of autism and PDD-NOS symptoms on challenging behaviors in adults with intellectual disabilities. Journal of Developmental and Physical Disabilities, 20, 41–51.

Mawhood, L., Howlin, P., & Rutter, M. (2000). Autism and developmental receptive language disorder – a comparative follow-up in early adult life. I: Cognitive and language outcomes. Journal of Child Psychology and Psychiatry, 41(5), 547–559.

Mullen, E. M. (1995). Mullen Scales of Early Learning. Circle Pines, MN: American Guidance Service.

Norrelgen, F., Fernell, E., Eriksson, M., Hedvall, Å., Persson, C., Sjölin, M., … Kjellmer, L. (2015). Children with autism spectrum disorders who do not develop phrase speech in the preschool years. Autism, 19(8), 934–943. https://doi.org/10.1177/1362361314556782.

Paul, R., Campbell, D., Gilbert, K., & Tsiouri, I. (2013). Comparing spoken language treatments for minimally verbal preschoolers with autism spectrum disorders. Journal of Autism and Developmental Disorders, 43, 418–431.

Paul, A., Sharda, M., Menon, S., Arora, I., Kansal, N., Arora, K., & Singh, N. C. (2015). The effect of sung speech on socio-communicative responsiveness in children with autism spectrum disorders. Frontiers in Human Neuroscience, 9, 555. https://doi.org/10.3389/fnhum.2015.00555.

Pickett, E., Pullara, O., O'Grady, J., & Gordon, B. (2009). Speech acquisition in older nonverbal individuals with autism: A review of features, methods, and prognosis. Cog. Behav. Neurol., 22, 1–21. https://doi.org/10.1097/WNN.0b013e318190d185.

Plesa-Skwerer, D., Jordan, S., Brukilacchio, B., & Tager-Flusberg, H. (2016). Comparing methods for assessing receptive language skills in minimally verbal children and adolescents with Autism Spectrum Disorders. Autism, 20(5), 591–604.

Reinhardt, V., Wetherby, A., Schatschneider, C., & Lord, C. (2015). Examination of sex differences in a large sample of young children with autism spectrum disorder and typical development. Journal of Autism and Developmental Disorders, 45, 697–706. https://doi.org/10.1007/s10803-014-2223-6.

Rice, M. (2016). Specific Language Impairment, Nonverbal IQ, Attention-Deficit/Hyperactivity Disorder, Autism Spectrum Disorder, Cochlear Implants, Bilingualism, and Dialectal Variants: Defining the Boundaries, Clarifying Clinical Conditions, and Sorting Out Causes. Journal of Speech, Language, and Hearing Research, 59, 122–132. https://doi.org/10.1044/2015_JSLHR-L-15-0255.

Rose, V., Trembath, D., Keen, D., & Paynter, J. (2016). The proportion of minimally verbal children with autism spectrum disorder in a community-based early intervention programme. Journal of Intellectual Disability Research, 60(5), 464–477.

Rogers, S., Hayden, D., Hepburn, S., Charlifue-Smith, R., Hall, T., & Hayes, A. (2006). Teaching young nonverbal children with autism useful speech: A pilot study of the Denver Model and PROMPT interventions. Journal of Autism and Developmental Disorders, 36, 1007–1024.

Schopler, E., Reichler, R. J., & Rochen-Renner, B. (1988). Childhood Autism Rating Scale. Los Angeles: Western Psychological Services.

Smith, V., Mirenda, P., & Zaidman-Zait, A. (2007). Predictors of expressive vocabulary growth in children with autism. Journal of Speech, Language, and Hearing Research, 50, 149–160.

StataCorp. (2015). Stata Statistical Software: Release 14. College Station, TX: StataCorp LP.

Tager-Flusberg, H., & Kasari, C. (2013). Minimally verbal school-aged children with autism spectrum disorder: The neglected end of the spectrum. Autism Research, 6(6), 468–478.

Tager-Flusberg, H., Plesa-Skwerer, D., Joseph, R., Brukilacchio, B., Decker, J., Eggleston, B., … Yoder, A. (2016). Conducting research with minimally verbal participants with autism spectrum disorder. Autism, 21, 1–10. https://doi.org/10.1177/1362361316654605.

Thurm, A., Lord, C., Lee, L.-C., & Newschaffer, C. (2007). Predictors of language acquisition in preschool children with autism spectrum disorders. Journal of Autism and Developmental Disorders, 37, 1721–1734. https://doi.org/10.1007/s10803-006-0300-1.

Venter, A., Lord, C., & Schopler, E. (1992). A follow-up study of high-functioning autistic children. Journal of Child Psychology and Psychiatry, 33(3), 489–507.

Wan, Y., & Schlaug, G. (2010). Neural pathways for language in autism: The potential for music-based treatments. Future Neurology, 5(6), 797–805.

Wan, C., Bazen, L., Baars, R., Libenson, A., Zipse, L., Zuk, J. ... G. Schlaug (2011) Auditory-motor mapping training as an intervention to facilitate speech output in non-verbal children with autism: A proof of concept study. PLoS One 6(9), e2550. doi:https://doi.org/10.1371/journal.pone.0025505, e25505

Wetherby, A., Watt, N., Morgan, L., & Shumway, S. (2007). Social communication profiles of children with autism spectrum disorders late in the second year of life. Journal of Autism and Developmental Disorders, 37, 960–975. https://doi.org/10.1007/s10803-006-0237-4.

White, I., & Thompson, S. (2005). Adjusting for partially missing baseline measurements in randomized trials. Statistics in Medicine, 24, 993–1007.

Yoder, P., & Stone, W. (2006). A randomized comparison of the effect of two prelinguistic interventions on the acquisition of spoken communication in preschoolers with ASD. Journal of Speech, Language, and Hearing Research, 49, 698–711.

Yoder, P., Watson, L., & Lambert, W. (2015). Value-added predictors of expressive and receptive language growth in initially nonverbal preschoolers with autism spectrum disorders. Journal of Autism and Developmental Disorders, 45, 1254–1270.