From intuition to intervention: developing an intonation-based treatment for autism

Karen Chenausky and Gottfried Schlaug

Music, Neuroimaging, and Stroke Recovery Laboratory, Dept. of Neurology, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, Massachusetts

Abstract

Autism affects ~ 1.5% of children under age 8; its core symptoms include impairment in social-communicative functioning and repetitive behaviors/restricted interests. Music-based interventions have been considered one modality through which to treat autism. This report discusses considerations to take into account when developing a music-based intervention for a core symptom of autism. Treatment modality must be matched to symptom both clinically and theoretically, the behavior to be treated must be carefully defined and assessed, and outcome measures must be capable of showing improvement in that behavior over the course of the study. Fidelity assessment and rater blinding reduce experimenter bias. High inter-rater reliability for perceptually determined outcome measures helps obtain accurate estimates of treatment response. Later stages of testing compare the experimental intervention to matched control treatments or other validated therapies, isolating the intervention's "active ingredients." Such systematic investigation of a new music-based intervention can provide information of different types, ranging from an assessment of whether the intervention has any effect at all to an assessment of its outcomes and risks in uncontrolled community settings. Findings ultimately compose the evidence base that clinicians and families can use to decide the most effective way of addressing symptoms of autism for particular children.

Keywords

autism; intonation; spoken language; developmental disorders

Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental disorder affecting approximately 1.5% of children.¹ It is defined by (1) the presence of repetitive behaviors/restricted interests, and (2) persistent deficits in social communication.²

The focus of the present work is to discuss how a music-based behavioral intervention for a core symptom of or a behavioral problem associated with ASD can be developed. Our

Address for correspondence: Karen V. Chenausky, Ph.D., CCC-SLP, Dept. of Neurology, Beth Israel Deaconess Medical Center, 330 Brookline Avenue, Palmer 127, Boston, MA 02215. kvchenau@bidmc.harvard.edu.

Competing interests

The authors declare no competing interests.

particular emphasis here will be using the development and testing of auditory–motor mapping training (AMMT), a novel intonation-based, spoken-language therapy for minimally verbal children with autism, to illustrate considerations to take into account when defining and testing behavioral treatments for core symptoms of ASD. It is not our purpose here to exhaustively enumerate the features of empirically based behavioral treatments for ASD, nor detailed models of systematic validation and dissemination of such interventions. For more comprehensive tutorials, see Refs. 3, 4, and 5. The underlying motivation for this article can be summarized by a quote from Ami Klin, Ph.D., in his keynote address at the 2017 International Society for Autism Research meeting: “We are clinicians who want to elevate our clinical feelings {about the therapeutic effect of a musical intervention} to the level of science” (addition ours).

Here, we discuss issues to consider at each stage of the development of a music-based intervention: (1) motivating the need for an intervention and matching the intervention to the behavior, (2) initial systematic application of the intervention, (3) replication and extension of initial findings, and (4) formally assessing efficacy in the lab and effectiveness in the community. Stages (1), (2), and (3) will be illustrated with examples and results from the ongoing development of AMMT. These stages, and some of their interactions, are diagrammed in Figure 1.

Engaging clinical intuition

The goal of intervention is to increase desired or functional behavior, improving the lives and outcomes of individuals affected by disorders. Clinical intuition provides the initial motivation for employing a specific intervention to modify a specific behavior. Thorough knowledge of the disorder whose symptoms will be treated must be combined with structured, informed observation of the target behavior and with clinical judgment that a particular type of intervention will improve it. This implies three basic steps at the very first stage of intervention development: (1) carefully characterize the disorder, its symptoms, and behaviors; (2) precisely define the target behavior or symptom; and (3) employ sensitive outcome measures that can track changes in symptoms or behaviors and demonstrate response to an intervention.

Careful documentation of diagnosis is vital in ASD research. Best clinical judgment from an experienced clinician should be employed, informed by the use of at least one gold-standard diagnostic instrument. The Autism Diagnostic Observation Schedule⁶ (ADOS) is an assessment given to the participant, while the Autism Diagnostic Interview-Revised⁷ (ADI-R) is a structured interview with a caregiver. Meeting criteria on both instruments provides convergent validity and increases diagnostic certainty. Both instruments require examiners to achieve a specified level of reliability on administration and scoring before they are qualified to use these instruments for research purposes. One advantage of using them, especially if participants are limited to, for example, ADOS Module 1 (for children with at most single words), is to reduce phenotypic heterogeneity, making it more likely that a between-group treatment effect will show up if the treatment is indeed effective.

Careful baseline characterization of study participants can also greatly aid identification of factors predicting response to treatment. This is true in two respects: first, because without an understanding of a learner's abilities and challenges, it is impossible to select what areas treatment should focus on, what supports may be necessary, or even to document progress.⁸ Second, even within a group of children who have been identified as being appropriate for a particular treatment, there is still some degree of heterogeneity. Since these individual differences may be related to treatment outcomes, it is important to capture this information in order to refine the treatment or better identify who may benefit from it.

Equal thought should be given to selection of a target behavior. Kazdin⁹ suggests targeting behavior that is disturbing, disruptive, or detrimental. Many symptoms of autism uncontroversially fit these criteria; however, mild repetitive behaviors may not. Behaviors such as eye contact are culturally and situationally dependent, calling for caution when working with individuals on the milder end of the spectrum or from other cultures. Regardless of the choice of treatment target, it is important to use objective, clear, and complete criteria to define a measurable behavior.¹⁰

Finally, outcome measures must have functional significance and should be carefully selected using quantifiable parameters, for example from standardized tests of IQ, socialization, or language. However, many researchers prefer to define their own measures, especially if none exists that covers the behavior in question. Regardless of whether an outcome measure is standardized, it is important that it possess good psychometric qualities.¹¹ For example, a good outcome measure will have high test-retest reliability (so that repeated baseline assessments will all produce similar scores, for example) and yet be sensitive to change (so that improvements can be detected). Carefully operationalizing definitional criteria; demonstrating high inter-rater identification reliability; and blinding raters to treatment group, time point, or other relevant variables are all necessary. Such procedures minimize experimenter bias, increase concept validity, and provide higher levels of confidence that any treatment effect is genuine.

A final issue at this stage is how large a change in the outcome measure defines "meaningful improvement" from a patient's or family's perspective. This is referred to as the *minimal clinically important difference*¹² (MCID). There are no hard-and-fast rules on what determines an MCID; it must be defined in the context of the disorder.¹³ The MCID should be based on a reliable and valid outcome measure and be larger than measurement error on that measure.¹⁴ Note that the MCID is not always constant across all levels of severity of a disorder:¹⁵ a one-point difference may represent a larger degree of improvement for severely-affected participants than for mildly-affected participants.

A recent study¹⁶ illustrates the importance of carefully defining the MCID. This randomized clinical trial (RCT) tested improvisational music therapy for young children with ASD, finding no significant between-group difference on the ADOS Social Affect subscore after 5 months. However, the ADOS was not designed to measure change in core symptoms of ASD over time.¹⁷ Cognizant of the critical need for outcome measures that are sensitive to improvement in social communication, a novel instrument, the Brief Observation of Social Communication Change¹⁸ (BOSCC) was developed. The BOSCC is appropriate for children

with minimal spontaneous language (appropriate for the ADOS modules employed by Bieleninik¹⁶) and can be used to score a 10-min free-play interaction. Two intervention studies, not employing music-based treatments, highlight the BOSCC's sensitivity to change. Kitzerow¹⁸ showed a significant decrease in BOSCC scores for 21 participants after one year of treatment (-5.6 points, $P < 0.001$, Cohen's $d = 0.63$), while ADOS severity scores only showed a trend for change (-0.61 , $P = 0.061$, $d = 0.42$). Similarly, Grzadzinski¹⁹ followed 50 children from a variety of treatment studies over a mean of 5.9 months. BOSCC scores significantly decreased (-2.53 , $P < 0.05$, $d = 0.26$), but ADOS social affect scores did not (-0.29 , $P > 0.05$, $d = 0.15$). The fact that the intervention of Bieleninik¹⁶ did not meet their MCID may therefore reflect a problem not with the intervention but with the outcome measure.

Motivation for AMMT

We wished to develop an intervention for spoken language in ASD. Approximately 25% of children with ASD remain minimally verbal (MV) past age 5: they use fewer than 20 words to communicate and do not combine words (e.g., "more juice").²⁰ MV status is associated with challenging behaviors such as aggression and self-injury.^{21,22} However, language level independently predicts positive long-term outcomes in ASD.²³ There is thus an urgent need for evidence-based interventions to teach children with MV ASD even a few words with which they can express their needs and wants. While developing useful speech by age 5 significantly predicts long-term language and communication outcomes for children with ASD,²⁴ the reality is that spoken language may not be feasible or achievable for all children with ASD who are at risk for remaining MV. However, communicating via speech maximizes the number of communication partners that a child has access to and is one way of improving communicative competence or the ability to express themselves to different listeners and in different contexts.²⁵ Thus, we have developed a speech-related intervention using intonation and hand-tapping on drums as one way of facilitating the engagement of the child and the mapping of sounds to actions.

Theoretical motivation for employing an intonation-based intervention for children with MV ASD comes from three lines of research. First, children with ASD often enjoy the motor behavior that goes with music^{26,27} and this can be capitalized on in treatment. Numerous studies have shown that mapping sounds with hand actions links perceptual and motor regions in the brain,²⁸ regions which may be dysfunctional in ASD.^{29–36}

Second, relative to typically developing peers, children with ASD show differences in brain regions subserving language. The right hemisphere structures involved with vocal communication may be more functionally significant for them. Homotopic right hemisphere structures have been found to be more sensitive to stimuli with large pitch variations. For example, Abrams³⁸ showed that the right hemisphere is more sensitive than the left to coding the envelope (the slow temporal features) of speech and may integrate information over longer time windows than corresponding left-hemisphere structures. Even very young children with ASD show anatomical and functional reversal of the usual left-right asymmetry in language networks in the inferior frontal gyrus.^{34,39–44} This asymmetry is supported by preliminary data showing volume and asymmetry abnormalities of the arcuate

fasciculus⁴⁵ (AF), a major fiber tract involved in language processing, in children with MV ASD.⁴² The AF connects regions that are important for mapping oral-motor actions onto auditory stimuli and for motor plan selection in response to auditory feedback and may play a critical role in the communication disorder in MV ASD. Children with ASD show micro- and macrostructural abnormalities in the AF, along with underconnectivity between frontal and posterior temporoparietal brain regions.^{35,36}

The third line of research suggesting that intonation-based interventions may be particularly efficacious for children with ASD comes from stroke rehabilitation literature. Melodic intonation therapy (MIT), an intonation-based treatment, has improved language fluency in left-hemisphere stroke patients.^{46,47} Taking all three lines together, we hypothesized that AMMT, adapted from MIT, would train the association between sounds and articulatory actions using bimanual movements performed simultaneously with intoned (sung) words and phrases. Synchronous drumming and singing might engage and stimulate the partially shared network of hand and articulatory actions^{28,48–49} that may be dysfunctional in children with MV ASD.^{29–33}

We defined “minimally verbal” as an expressive vocabulary (words used and understood) of fewer than 20 words.²⁰ To be eligible for the studies, children must have had a community diagnosis of autism, verified by a score of > 30 on the Childhood Autism Rating Scale⁵⁰ or > 12 on the ADOS. Expressive vocabulary was verified by the number of different words produced during the ADOS or a 21-min language sample at baseline. Because imitation ability has been shown to predict expressive language,^{51,52} participants must also be able to imitate at least two speech sounds. Children with echolalia were not explicitly excluded; however, potential participants with long echolalic phrases generally failed to meet criteria on other grounds (i.e., they had too large a productive vocabulary). Finally, children with diagnoses of sensorineural disorders other than ASD (e.g., deafness, Down’s Syndrome) were excluded.

Our main outcome measure was % syllables approximately correct, over a set of 30 bisyllabic words and phrases used in both AMMT and our control treatment. Though acoustic and kinematic measures of speech are objective and available, we used a perceptually based measure because it is both clinically meaningful and ecologically valid as a proxy for speech intelligibility.^{53–55} Adding to the ecological validity, our scoring rubric took into account that children were not expected to produce the stimuli with adult-like skill. A syllable was considered approximately correct if (1) the consonant produced shared two of three phonetic features (voicing, place of articulation, manner of articulation) with the target and (2) the vowel was within the same class as the target, sharing two features (tongue height and backness, which refer to dorsal/ventral and anterior/posterior position within the mouth, respectively) with the target.

Stimuli were selected to represent common objects (“bubbles”), actions (“all done”), and people (“daddy”) relevant to children’s lives. Twenty-eight of the words appear in the MacArthur-Bates Communication Development Inventory,⁵⁶ and the remaining ones are in the top 1008 most frequently used words from the Corpus of Contemporary American English.⁵⁷ Stimuli were represented by iconic images printed on laminated cards, which

were presented to the child at the start of each trial and removed at its end. Occasionally a real object (“water,” “bubbles,” “cookie”) was used if available, especially if the named object could be used as a reward. Photographs of the child’s parents were also used in place of iconic images in some cases.

In terms of phonological complexity, a range of phonemes and syllables structures was included. While the stimuli were heavily weighted toward early-appearing phonemes, such as /b/, /m/, and /d/, developmentally later-appearing consonants, such as /p/, /r/, and /ʃ/, were also included,⁵⁸ because participants varied in abilities, and it was important to avoid both floor and ceiling effects as much as possible. Similarly, syllable shapes included various combinations of vowels (V) and consonants (C): V, CV, VC, CVC, and CCVC, also spanning a range of complexity.

As it is currently implemented, AMMT focuses mainly on training children to more intelligibly produce stimuli; less emphasis is put on using the stimuli referentially, although children are encouraged to use stimuli such as “hello” and “bye-bye” at the beginning and end of sessions, and “more please” to request activities or rewards. Our view is that intelligible speech production is a necessary but not sufficient component for successful use of spoken language and is thus an important treatment focus. Since a range of skills is involved in spoken language development,²⁴ AMMT fits in with other targeted interventions that aim to teach pre-linguistic and communication skills.

Performing multiple baseline assessments before initiating treatment is a staple of single-case experimental designs⁷ and serves several functions. Multiple baseline assessments allow children to acclimate to the examiners and environment, establish a stable performance level, and provide an estimation of test/retest variability on the outcome measure prior to intervention. A child’s best baseline score (the session with the highest number of syllables approximately correct) was subtracted from their score after treatment to assess improvement. To minimize perceptual bias, we employed an explicit rubric for determining whether a child’s production sufficiently accurately approximated the adult target.^{59–62} Multiple raters, blinded to assessment time point, phonetically transcribed children’s responses during baseline and posttreatment assessments.

Adequate inter-rater reliability maximizes precision and statistical power and avoids obscuring a treatment effect.⁶³ Inter-rater scoring reliability was assessed using a set of 10% of assessment sessions across participants. An independent rater scored this set of sessions and calculated percent agreement (the number of items agreed upon divided by the total number of items), a straightforwardly interpretable quantity commonly used in transcription analyses.⁶⁴ Agreement was 68.0% on percent (%) syllables approximately correct, which compares favorably to previous figures for infant babbles (61% agreement, averaged over consonants and vowels).⁶⁵ Because percent agreement does not take into account the agreement between raters that would be due solely to chance,⁶³ we also used Cohen’s κ .⁶⁶ For syllables approximately correct, $\kappa = 0.547$, $P < 0.0005$. There are no hard-and-fast rules for determining acceptable values of κ , but guidelines^{67,68} classify this as at least “moderate” or “good.”

Systematic application of the intervention

The first step in empirical validation is to systematically apply the intervention to one or a small number of participants,³ establishing initial evidence for a treatment effect. Participants selected at this step are those most likely to benefit from the treatment. Experimental designs include both single-case and uncontrolled group studies.⁹ If an effect is present, it will result in a large between-group difference that will clearly indicate whether the intervention's success merits taking it to the next step. That is, the goal of this first step is to collect the information needed to proceed to the next one, as well as information related to how the intervention may work.⁵

To apply an intervention successfully to multiple participants, therapeutic procedures must be formalized so that they can be reproduced faithfully by different practitioners. Manualizing a therapy requires piloting a specific set of procedures and revising them as needed until a consistently effective set is achieved. These procedures can then be used to assess treatment fidelity across multiple therapists and participants.^{3,4}

Other steps in the initial stages of developing an intervention may include feasibility testing, which can assess the practicalities of providing the treatment in community settings.^{3,5} The feasibility process may involve activities such as finding independent research whose results bear on the theoretical aspects of the intervention, using that research to formulate hypotheses about the mechanism of action of the intervention that can be tested in later phases, and iteratively refining the intervention to maximize positive effects and make it as straightforward as possible to implement.³

Systematizing AMMT

In AMMT, stimuli are intoned at a rate of one syllable/s, using two pitches that follow the words' natural prosodic contour. At the same time, therapist and child tap electronic drums tuned to the same pitches, one tap per syllable. A simple, straightforward relationship between musical notes and prosodic structure was chosen because music and language understanding are related to the level of language disorder.⁶⁹

The following steps are used in AMMT:

1. Listening: Therapist introduces target phrase by showing a picture and using it in a semantic context: "It's fun to blow *bubbles*." Therapist produces target.
2. Unison: "Let's say it together: '*bubbles*.'" Therapist produces target with child.
3. Unison fade: "Again: '*bu...*'" Therapist produces initial portion of the target with child; child continues alone.
4. Imitation: (4a) "My turn: '*bubbles*.'" Therapist produces target. (4b) "Your turn: ..." Child imitates target.
5. Cloze: "Last time: It's fun to blow..." Therapist presents target in the same semantic context; child completes it, producing target independently.

These steps form a hierarchy. The listening step imposes the fewest demands. Unison production provides maximal support; unison fade less support, immediate repetition still less, and independent production the least. This temporal delay hierarchy is a validated strategy that provides graduated opportunities, giving the child more and more responsibility for retrieving and assembling the motor plan for each word.⁷⁰ Each stimulus was practiced five times, as described above, then another stimulus was practiced. This procedure was repeated until all 15 trained stimuli were practiced in each treatment session. This allowed for both massed practice (many repetitions in a row) and distributed practice (spreading trials over a longer period of time)—principles of motor learning that are important for teaching speech movements.⁷¹

In addition to employing scaffolding hierarchies, massed practice, and distributed practice, we also employed a range of behavioral management techniques as needed. For example, children were positively reinforced for attempting responses by receiving frequent social praise and tangible rewards (edibles, preferred objects, or preferred activities) on variable-interval schedules. Planned ignoring was employed in most cases of mild disruptive behavior. Challenging behaviors were rare and generally addressed by withdrawal of attention, combined with positive reinforcement for appropriate behaviors (such as attempting a vocalization). Subsequently, the child was then prompted to perform an action that had a high probability of success, such as “touch your head” or “clap your hands.” In addition to providing the child an opportunity to experience success and earn a reward, this technique maintained the reciprocal contingency with the adult (“you do something for me, I do something for you”). After the child was rewarded for compliance, the session could either continue or be terminated, based on the therapist’s judgment.

Assessment (probe) sessions were conducted after the 10th treatment session and after every five sessions thereafter. Structure of treatment and assessment sessions was identical (items were presented randomly; all five steps for one item were completed before moving to the next item) except that an additional 15 untrained stimuli were presented during assessment sessions but not practiced during treatment. The purpose of including these stimuli was to assess the degree of generalization that children showed in their developing speech production skills. Assessments were administered by the child’s therapist but, as discussed above, were scored by raters blind to session date.

Another outcome of treatment manualization is a set of hypothesized “active ingredients,”⁴ or features of the intervention considered necessary for its effect. For AMMT, these are (1) intoning the stimuli and (2) simultaneous drumming. Rhythmic drum-tapping may activate the auditory–motor network mapping sounds to actions and facilitate speech production by entraining the vocal system and recapitulating the “babbling and banging” stage of speech development.^{72,73}

AMMT was fine-tuned in three pilot participants with varying degrees of spoken-language impairment. We established the initial efficacy of AMMT in a proof-of-concept study using within-subject analyses, reporting results for six participants with MV ASD (1 F) aged 5;9 to 8;9 who received 40 treatment sessions. Participant characteristics are listed by study in successive rows of Table 1. In this phase of AMMT’s development, each participant

functioned as his or her own control, with the goal of identifying whether AMMT could consistently produce improvement in a larger group of participants. Best Baseline scores averaged 15.0% syllables approximately correct (range 0.0–28.3, SD 12.4); posttreatment scores averaged 51.0% (range 36.7–85.0, SD 17.6). Participants improved by an average of 37.0% syllables approximately correct (SD 11.6). A paired-samples *t*-test was significant ($P = 0.001$), and binomial tests of significance revealed that baseline scores were below the 95% confidence interval for posttreatment scores for all participants.⁵⁹ The mean baseline score of 15% syllables approximately correct corresponds to nine out of 60 syllables approximated, and the mean posttreatment score of 51% syllables approximately correct corresponds to ~ 30 syllables approximated. The increase from fewer than 10 to over half of syllables in trained and untrained stimuli produced approximately correctly was deemed highly clinically significant.

Replication of initial findings

Given these promising initial findings, the next step was to replicate these results in a larger group of 17 participants (2 F; aged 3;5–9;8; second row of Table 1).

A methodological issue arose at this point. Forty sessions of treatment, plus baseline and planned assessment sessions, posed a significant logistical burden for families. Because the magnitude of change after 25 treatment sessions for the proof-of-concept participants (32.8%, SD 6.9) represented 88.6% of the change after 40 sessions and was still statistically significant ($P > 0.0005$), we shortened the course of treatment to 25 sessions. Session length remained constant at approximately 45 minutes.

To establish treatment fidelity for both treatments (i.e., to train new therapists), each new therapist first observed sessions with an experienced therapist, then administered a series of sessions with the experienced therapist sitting in and correcting any mistakes. After the first session where no corrections were needed, the therapist was allowed to treat independently, and performance was monitored periodically thereafter (approximately one of every 10 sessions). To assess treatment fidelity for AMMT and to verify that SRT did not contain the hypothesized active ingredients, an independent rater viewed video of a total of 26 sessions (11%). Stimuli were intoned and drums used on all AMMT trials and no SRT trials. As previously reported,⁶⁰ 0.6% of trials included repeated steps and only 0.1% had omitted steps.

In the 17 replication participants, mean % syllables approximately correct at baseline was 29.3% (range 3.4–63.4, SD 17.1); after 25 sessions (P25), it was 44.5% (range 5.0–93.4, SD 29.9). The average improvement of 15.2% syllables approximately correct (SD 21.7) was statistically significant according to paired *t*-tests ($P = 0.01$). A repeated-measures ANOVA with time (baseline vs. P25) as a within-subjects factor and group (proof of concept vs. replication) as a between-subjects factor revealed a main effect of time ($F(1,21) = 25.566$, $P < 0.0005$), indicating that the combined group of 23 AMMT participants improved on average. There was no significant main effect of group ($F(1,21) = 0.241$, $P = 0.628$), indicating that the proof-of-concept and replication groups were not consistently different over time. Finally, there was no significant time $\times$ group interaction ($F(1,21) = 2.885$, $P =$

0.104), meaning that the proof of concept and replication groups did not have different trajectories over time. Because the two groups were statistically equivalent, they were combined into one analysis. Figure 2 shows the improvement of the proof-of-concept and replication groups over time.

Combining the six proof-of-concept participants with the 17 replication participants (23 total), mean % syllables approximately correct at baseline was 26.1% (range 0–63.4, SD 16.5); at P25, the mean was 45.5% (range 5–93.4, SD 25.9). The average improvement of 19.4% syllables approximately correct (SD 19.9) was statistically significant (paired *t*-tests, $P = 0.001$). Note that the amount of change in the larger group of 23 AMMT participants at P25 was approximately 60% of the change observed in the original group of six participants over the same time period. However, the mean improvement from ~ 15 of 60 syllables produced approximately correctly at baseline to ~ 27 produced approximately correctly after 25 sessions (almost double the number at baseline) was still considered clinically significant. This reduction in magnitude of change as more participants received the therapy was not unexpected: the strategy of choosing initial participants who were likely to show a strong treatment effect meant that subsequent tests of the treatment may not result in as large a change.

A further test of an intervention is to compare it to another treatment. As an initial investigation of the two hypothesized active ingredients of AMMT (simultaneous intonation and drumming), we compared AMMT to a control treatment, speech repetition therapy (SRT). While lacking AMMT's active ingredients, SRT included the other aspects of empirically validated treatments, such as a scaffolding hierarchy and opportunities for both blocked and distributed practice. Both treatments employed the same set of bisyllabic stimuli and provided the same dose of therapy. Seven AMMT participants from the replication group of 17 were matched to seven SRT participants on the basis of chronological age, mental age, and phonemic repetition ability at baseline.

A repeated-measures ANOVA comparing baseline and P25 performance revealed a significant main effect of time ($F(1,12) = 23.508$, $P < 0.0005$), indicating that on average the entire group of 14 children improved from baseline to post 25 sessions. Mean change was + 15.5%. There was no significant main effect of treatment, so the AMMT and SRT groups were not consistently different across all time points. There was, however, a significant time x treatment interaction ($F(1,12) = 14.467$, $P = 0.003$), indicating that the groups showed different trajectories over time. As previously reported,⁵² the AMMT group improved by a mean of 29.0% syllables approximately correct from baseline to P25, while the SRT group improved by 3.6%.

In addition to comparing the AMMT and SRT groups' performance immediately after treatment, we also examined whether changes would be maintained at assessments 4 or 8 weeks posttreatment. These results are reported for the first time here. Ten AMMT participants and seven SRT participants who received 25 treatment sessions returned for 4- and 8-week posttreatment follow-ups. (The remaining 13 AMMT participants either received more than 25 treatment sessions or lacked 4- or 8-week follow-up data.)

A repeated-measures ANOVA comparing % syllables approximately correct for the 17 25-session participants with follow-up data showed a main effect of time ($F(3,45) = 16.250$, $P < 0.0005$). On average, these 17 participants improved from baseline to P25 + 8 weeks. The difference between baseline and P25 scores (+14.0% syllables approximately correct) was significant ($P < 0.001$). Neither the difference between P25 and P25 + 4 weeks scores (+ 1.1% syllables approximately correct) nor between P25 + 4 weeks and P25+8 weeks (– 1.4% syllables approximately correct) were significant (both $P = 1.0$), indicating that the improvements at P25 persisted into the posttreatment period. There was also a significant main effect of stimuli, such that trained stimuli were produced approximately correctly more often (+ 6.7%) than untrained stimuli. There was no significant main effect of group ($F(1,15) = 0.569$, $P = 0.462$), indicating that the AMMT and SRT groups were not consistently different from each other across time points. However, the significant time $\times$ group interaction was maintained ($F(3,45) = 8.948$, $P < 0.0005$). There were no other significant main effects or interactions. The time $\times$ group figures are tabulated in Table 2.

Taken together, the results from the larger group of 23 AMMT participants and the matched AMMT and SRT groups continued to demonstrate a treatment effect. Though the magnitude of improvement decreased slightly and clinically insignificantly from the proof-of-concept stage to the replication stage, the difference between the mean improvement in the matched groups still corresponded to a very large effect size (Cohen's $d = 3.0$)⁶⁰ and was thus considered a sufficiently strong result to warrant continuing to the next step, a randomized clinical trial (RCT).

Ongoing and future work: clinical trials and hypothesis testing

The RCT is the gold-standard method to demonstrate clinical effectiveness,³ because randomization creates groups differing only by treatment. RCTs are challenging to carry out, requiring large numbers of participants and often cessation of other therapies. Thus, ethical considerations come into play. Clinical effectiveness trials, where two treatments are compared to each other, can avoid this problem.

Another challenge in conducting RCTs is that participants who do not improve are much less motivated to remain in the study, leading to skewed results. Adaptive trial designs, such as the sequential multiple assignment randomized trial^{74,75}(SMART) address this issue by allowing for adjustments when participants do not show a pre-specified degree of improvement after a certain amount of time. In SMART trials, children are initially randomized to a treatment. Children who do not improve after (say) 12 weeks are then randomized to either (1) an increased dose of the same treatment or (2) addition of another treatment to the first.⁷⁵ There are, in addition, experimental designs other than clinical trials⁷ that can be used to establish whether an intervention causes an improvement. For example, reversal, multiple-baseline, and changing-criterion single-subject designs are non-RCT methods of ascertaining a causal effect of an intervention.

When conducting a trial, however, prospective power analyses⁷⁶ help determine how many participants are needed to show a specific effect and to minimize the risk of false negatives. There are several methods for prospectively determining effect sizes, including referring to

meta-analyses of studies that examine the relevant treatment effect; estimating the effect size based on any previous research (including one's own prior studies); and constructing a table to explore trade-offs between number of participants, effect size, and level of statistical power.⁷⁶ A variety of power-analysis calculators is available online (e.g., <http://www.biostat.ucsf.edu/samplesize.html>).

Currently, we are conducting an RCT comparing AMMT and SRT, using the same inclusion/exclusion criteria and treatment procedures as in earlier stages. Using the standard deviation for % syllables approximately correct that was observed in the initial six participants, we anticipated an effect size of 2.25. To detect this with 90% power requires 10 participants per group. Of course, degree of improvement is not the only aspect of AMMT and SRT that can be tested in the context of an RCT. Other hypotheses about AMMT can also be addressed; for example, that music-based treatment is more enjoyable or motivating for many children with MV ASD. As mentioned by one of the reviewers, this could be accomplished by behavioral coding of the children to determine average demeanor or the rate of escape-maintained behavior during a session, for example.

Designing an intervention does not stop after an RCT, however. The intervention must transition from a controlled laboratory environment to the real world—where children who may not meet the original inclusion criteria receive multiple treatments simultaneously from clinicians who are necessarily generalists. Future testing of AMMT could involve conducting so-called pragmatic trials⁷⁷ to assess its effectiveness with extended populations and multiple sites. In pragmatic trials, community therapists would be trained in the basic principles of AMMT, and the treatment provided in classrooms or community therapy centers.

Another paradigm for post-RCT study of interventions is comparative effectiveness research⁷⁸ (CER). CER trials directly compare the outcomes, effectiveness, risks, and benefits of two or more treatments and provide information that clinicians can apply to typical patients.⁷⁸ However, because differences between proven therapies are expected to be very small, CER trials require longer follow-up periods and larger numbers of participants. In addition, CER data is often challenging to interpret because of higher rates of missing data and confounds, which all affect treatments delivered in actual community settings.

Thus, some researchers provide experimental therapies as adjuncts to treatment as usual (TAU). In the case of treatment for children with MV ASD, TAU generally consists of 20–30 hours per week of applied behavior analysis-based treatments, structured into 30-min periods of 1:1 or 1:2 behavioral work with a therapist, interspersed with snack, lunch, and recess periods. In one such study,⁷⁹ daily half-hour sessions of an experimental treatment were added to children's TAU regimens. AMMT could similarly be tested as an adjunct treatment targeting spoken language; in this case, data would have to be collected about the volume and types of other treatments that participants receive.

Finally, there are a variety of other aspects of intervention development that are worth mentioning, though they are beyond the scope of this article to discuss in detail. Feasibility testing assesses whether the capacity of a particular location is conducive to carrying out an

RCT⁸⁰ and is important when considering implementing a treatment internationally. Generalizability⁸¹ is a way of determining whether the findings of a particular RCT can be extended to similar populations; for example, whether the effects of a music-based intervention for social interaction, originally designed for children with autism, might generalize to children with social communication disorder or language disorder. Translation, implementation, and maintenance are later stages of intervention development, in which a proven intervention is embedded into ongoing clinical practice and adopted as part of the standard of care.⁸²

Conclusions

Different considerations apply at different stages in an intervention's "life span." Treatment modality must be motivated clinically and theoretically. Music-based interventions may provide a unique avenue for improving spoken language in children with MV ASD by engaging them in enjoyable turn-taking activities that capitalize on their strengths and involve mapping auditory to motor actions. Precise characterization of the target behavior and careful selection of outcome measures maximize the ability to detect a treatment effect. At all stages, blinding, fidelity assessment, and inter-rater reliability assessments reduce the effect of experimenter bias. During later stages, larger samples are needed to demonstrate a predetermined effect size. Final testing must occur in community settings, assessing an intervention's viability in real-world contexts. These procedures inform clinical decision making, from determining whether the intervention has any effect at all to assessing its outcomes and risks in uncontrolled community settings.

Acknowledgments

We thank our former speech-language pathology clinical fellows, postdoctoral fellows, and junior investigators, music therapists, and all research assistants and summer students for their hard work in the collection of these data and the conduct and design of these pilot studies, especially our esteemed colleague Andrea C. Norton. We are deeply grateful for the effort of the dedicated families who committed significant amounts of their time to making this work possible. K.C. and G.S. accept responsibility for data integrity. G.S. conceived and designed the original studies. K.C. and G.S. substantially contributed to acquisition and analysis of data, drafting and revising the manuscript, and approving the final version.

References

1. Christensen D, Baio J, Braun K, et al. Prevalence and characteristics of autism spectrum disorder among children aged 8 years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States. *MMWR Surveill Summ*. 2012; 65(SS-3):1–23. 2016. DOI: <http://dx.doi.org/10.15585/mmwr.ss6503a1>.
2. American Psychiatric Association. American psychiatric association task force on DSM-5: Diagnostic and statistical manual of mental disorders: DSM-5. Washington, DC: APA; 2013.
3. Smith T, Scahill L, Dawson G, et al. Designing research studies on psychosocial interventions in autism. *Journal of Autism and Developmental Disorders*. 2007; 37:354–366. [PubMed: 16897380]
4. Schreibman L, Dawson G, Stahmer A, et al. Naturalistic developmental behavioral interventions: Empirically validated treatments for autism spectrum disorder. *Journal of Autism and Developmental Disorders*. 2015; 45:2411–2428. [PubMed: 25737021]
5. Onken L, Carroll K, Shoham V, et al. Reenvisioning clinical science: Unifying the discipline to improve public health. *Clinical Psychological Science*. 2014; 2(1):22–34. [PubMed: 25821658]

6. Lord C, Risi S, Lambrecht L, et al. The Autism Diagnostic Observation Schedule - Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*. 2000; 30:205–223. [PubMed: 11055457]
7. Rutter M, LeCouteur A, Lord C. Autism Diagnostic Interview – Revised. Los Angeles: Western Psychological Services. 2008
8. Shane H, Weiss-Kapp S. Visual Language in Autism. San Diego: Plural Publishing; 2008.
9. Kazdin A. Single-Case Research Designs: Methods for Clinical and Applied Settings. New York: Oxford University Press; 1982.
10. Mayer G, Sulzer-Azaroff B, Wallace M. Behavior Analysis for Lasting Change. Third Edition. Cornwall-on-Hudson, NY: Sloan Publishing; 2014.
11. Anagnostou E, Jones N, Huerta M, et al. Measuring social communication behaviors as a treatment endpoint in individuals with autism spectrum disorder. *Autism*. 2015; 19(5):622–636. [PubMed: 25096930]
12. Jaeschke R, Singer J, Guyatt G. Measurement of health status: Ascertaining the minimal clinically important difference. *Controlled Clinical Trials*. 1989; 10:407–415. [PubMed: 2691207]
13. Beaton D, Boers M, Wells G. Many faces of the minimal clinically important difference (MCID): A literature review and directions for future research. *Current Opinion in Rheumatology*. 2002; 14:109–114. [PubMed: 11845014]
14. Jacobson N, Roberts L, Berns S, et al. Methods for defining and determining the clinical significance of treatment effects: Description, application, alternatives. *J Consulting Clin Psychol*. 1999; 67:300–307.
15. Riddle D, Stratford P, Binkley J. Sensitivity to change of the Roland-Morris back pain questionnaire: Part 2. *Phys Ther*. 1998; 78:1197–1207. [PubMed: 9806624]
16. Bieleninik L, Geretsegger M, Mossler K, et al. Effects of improvisational music therapy vs enhanced Standard care on symptom severity among children with autism spectrum disorder: The TIME-A randomized clinical trial. 2017. *JAMA*. 2017; 318(6):525–535. DOI: 10.1001/jama.2017.9478 [PubMed: 28787504]
17. Gotham K, Pickles A, Lord C. Standardizing ADOS scores for a measure of severity in autism Spectrum disorders. *Journal of Autism and Developmental Disorders*. 2009; 39:693–705. DOI: 10.1007/s10803-008-0674-3 [PubMed: 19082876]
18. Kitzerow J, Teufel K, Wilker C, et al. Using the Brief Observation of Social Communication Change (BOSCC) to measure autism-specific development. *Autism Res*. 2015; 9:940–950. DOI: 10.1002/aur.1588 [PubMed: 26643669]
19. Grzadzinski R, Carr T, Colombi C, McGuire K, Dufek S, et al. Measuring changes in social communication behaviors: Preliminary development of the brief observation of social Communication change (BOSCC). *Journal of Autism and Developmental Disorders*. 2016; 46:2464–2479. DOI: 10.1007/s10803-016-2782-9 [PubMed: 27062034]
20. Kasari C, Brady N, Lord C, et al. Assessing the minimally verbal school-aged child with autism spectrum disorder. *Autism Research*. 2013; 6:479–493. [PubMed: 24353165]
21. Hartley S, Sikora D, McCoy R. Prevalence and risk factors of maladaptive behaviour in young children with autistic disorder. *J Intellect Disabil Res*. 2008; 52:819–829. DOI: 10.1111/j.1365-2788.2008.01065.x [PubMed: 18444989]
22. Dominick K, Davis N, Lainhart J, et al. Atypical behaviors in children with autism and children with a history of language impairment. *Res Dev Disabil*. 2007; 28:145–162. DOI: 10.1016/j.ridd.2006.02.003 [PubMed: 16581226]
23. Buschbacher P, Fox L. Understanding and intervening with the challenging behavior of young children with autism spectrum disorder. *Lang Speech Hear Serv Sch*. 2003; 34:217–227. DOI: 10.1044/0161-1461(2003/018) [PubMed: 27764323]
24. Hampton L, Kaiser A. Intervention effects on spoken-language outcomes for children with autism: A systematic review and meta-analysis. *Journal of Intellectual Disability Research*. 2016; 60(5): 444–463. [PubMed: 27120988]
25. Light J, McNaughton D. Communicative competence for individuals who require augmentative and alternative communication: A new definition for a new era of communication? *Augmentative and Alternative Communication*. 2014; 30:1–18.

26. Trevarthen C, Aitken K, Paoudi D. , et al. Children with Autism. London: Jessica Kingsley Publishers; 1996.
27. Heaton P, Hermelin B, Pring L. Autism and pitch processing: A precursor for savant musical ability? *Music Perception*. 1998; 15:291–305.
28. Lahav A, Saltzman E, Schlaug G. Action representation of sound: audiomotor recognition network while listening to newly acquired actions. *J Neurosci*. 2007; 27:308–314. [PubMed: 17215391]
29. Hadjikhani N. Mirror neuron system and autism. Carlisle PC, editor *Progress in Autism Research* Nova Science Publishing Inc; 2007. 151–166.
30. Iacoboni M, Dapretto M. The mirror neuron system and the consequences of its dysfunction. *Nat Rev Neurosci*. 2006; 7:942–951. [PubMed: 17115076]
31. Oberman L, Ramachandran V. The simulating social mind: The role of the mirror neuron system and simulation in the social and communicative deficits of autism spectrum disorders. *Psychol Bull*. 2007; 133:310–327. [PubMed: 17338602]
32. Oberman L, Ramachandran V, Pineda J. Modulation of mu suppression in children with autism spectrum disorders in response to familiar or unfamiliar stimuli: The mirror neuron hypothesis. *Neuropsychologia*. 2008; 46:1558–1565. [PubMed: 18304590]
33. Wan C, Schlaug G. Music making as a tool for promoting brain plasticity across the lifespan. *Neuroscientist*. 2010; 16(5):566–577. [PubMed: 20889966]
34. De Fosse L, Hodge S, Makris N, et al. Language association cortex asymmetry in autism and specific language impairment. *Ann Neurol*. 2004; 56:757–766. [PubMed: 15478219]
35. Just M, Cherkassky V, Keller T, et al. Cortical activation and synchronization during sentence comprehension in high-functioning autism: evidence of underconnectivity. *Brain*. 2004; 127:1811–1821. [PubMed: 15215213]
36. Stigler K, McDonald B, Anand A, et al. Structural and functional magnetic resonance imaging of autism spectrum disorders. *Brain Research*. 2011; 1380:146–161. [PubMed: 21130750]
37. Zatorre R, Gandour J. Neural specializations for speech and pitch: Moving beyond the dichotomies. Vol. 363. *Philosophical Transactions of the Royal Society of London*; 2007. 1087–1104.
38. Abrams D, Nicol T, Zecker S, et al. Right-hemisphere auditory cortex is dominant for coding syllable patterns in speech. *Journal of Neuroscience*. 2008; 28(15):3958–3965. [PubMed: 18400895]
39. Herbert M, Ziegler D, Deutsch C, et al. Brain asymmetries in autism and developmental language disorder: a nested whole-brain analysis. *Brain*. 2005; 128:213–226. [PubMed: 15563515]
40. Tager-Flusberg H, Lindgren K, Mody M. Structural and functional imaging research on language disorders: Specific language impairment and autism spectrum disorders. In: Wolf L, Schreiber H, Wasserstein J, editors *Adult learning disorders: Contemporary issues*. New York: Psychology Press; 2008. 127–157.
41. Kleinhans N, Muller RA, Cohen D, et al. Atypical functional lateralization of language in autism spectrum disorders. *Brain Research*. 2008; 1221:115–125. [PubMed: 18555209]
42. Fletcher P, Whitaker R, Tao R, et al. Microstructural connectivity of the arcuate fasciculus in adolescents with high-functioning autism. *Neuroimage*. 2010; 51:1117–1125. [PubMed: 20132894]
43. Knaus T, Silver A, Kennedy M, et al. Language laterality in autism spectrum disorder and typical controls: A functional, volumetric, and diffusion tensor MRI study. *Brain and Language*. 2010; 112:113–120. [PubMed: 20031197]
44. Redcay E, Courchesne E. Deviant fMRI patterns of brain activity to speech in 2–3 year-old children with autism spectrum disorder. *Biological Psychiatry*. 2008; 64(7):589–598. [PubMed: 18672231]
45. Wan C, Marchina S, Norton A, et al. Atypical hemispheric asymmetry in the arcuate fasciculus of completely nonverbal children with autism. *Ann NY Acad Sci*. 2012; 1252:332–337. [PubMed: 22524376]
46. Schlaug G, Marchina S, Norton A. From singing to speaking: Why patients with Broca's aphasia can sing and how that may lead to recovery of expressive language functions. *Music Percept*. 2008; 25:315–323. [PubMed: 21197418]

47. Wan C, Zheng X, Marchina S, et al. Intensive therapy induced contralateral white matter changes in chronic stroke patients with Broca's aphasia. *Brain and Language*. 2014; 136:1–7. [PubMed: 25041868]
48. Meister I, Boroojerdi B, Foltys H, et al. Motor cortex hand area and speech: implications for the development of language. *Neuropsychologia*. 2003; 41:401–406. [PubMed: 12559157]
49. Meister I, Buelte D, Staedtgen M, et al. The dorsal premotor cortex orchestrates concurrent speech and fingertapping movements. *European Journal of Neuroscience*. 2009; 29:2074–2082. [PubMed: 19453637]
50. Schopler E, Reichler R, Renner B, Rothen. *Childhood Autism Rating Scale*. Los Angeles: Western Psychological Services; 1988.
51. Smith V, Mirenda P, Zaidman-Zait A. Predictors of expressive vocabulary growth in children with autism. *Journal of Speech, Language, and Hearing Research*. 2007; 50:149–160.
52. Thurm A, Lord C, Lee LC, et al. Predictors of language acquisition in preschool children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*. 2007; 37:1721–1734. DOI: 10.1007/s10803-006-0300-1 [PubMed: 17180717]
53. Rogers S, Hayden D, Hepburn S, et al. Teaching young nonverbal children with autism useful speech: A pilot study of the Denver Model and PROMPT interventions. *J Autism Dev Disord*. 2006; 36:1007–1024. [PubMed: 16845576]
54. Yoder P, Watson L, Lambert W. Value-added predictors of expressive and receptive language growth in initially nonverbal preschoolers with autism spectrum disorders. *Journal of Autism and Developmental Disorders*. 2015; 45:1254–1270. [PubMed: 25344152]
55. Paul R, Campbell D, Gilbert K, et al. Comparing spoken language treatments for minimally verbal preschoolers with autism spectrum disorders. *Journal of Autism and Developmental Disorders*. 2013; 43:418–431. [PubMed: 22733301]
56. Fenson L, Dale P, Reznick J, et al. *The MacArthur Communicative Development Inventories: User's Guide and Technical Manual*. Baltimore, MD: Paul H. Brookes Publishing Co; 1993.
57. Davies M. The corpus of contemporary American English (COCA): 520 million words, 1990–present. 2008. Available online at: <http://corpus.byu.edu/coca/>
58. Shriberg L. Four new speech and prosody-voice measures for genetics research and other studies in developmental phonological disorders. *Journal of Speech and Hearing Research*. 1993; 36:105–140. [PubMed: 8450654]
59. Wan C, Bazen L, Baars R, et al. Auditory-motor mapping training as an intervention to facilitate speech output in non-verbal children with autism: A proof of concept study. *PLoS One*. 2011; 6:e25505.doi: 10.1371/journal.pone.0025505 [PubMed: 21980480]
60. Chenausky K, Norton A, Tager-Flusberg H, et al. Auditory Motor Mapping Training: Comparing the effects of a novel speech treatment to a control treatment for minimally verbal children with autism. *PLoS ONE*. 2016; 11(11):e0164930.doi: 10.1371/journal.pone.0164930 [PubMed: 27829034]
61. Chenausky K, Norton A, Schlaug G. Auditory-motor mapping training in a more verbal child with autism. *Frontiers in Human Neuroscience*. 2017; 11:426.doi: 10.3389/fnhum.2017.00426 [PubMed: 28928645]
62. Chenausky K, Kernbach J, Norton A, et al. White matter integrity and treatment-based change in speech performance in minimally verbal children with autism spectrum disorder. *Frontiers in Human Neuroscience*. 2017; 11 article 175. doi: 10.3389/fnhum.2017.00175
63. Hunt R. Percent agreement, Pearson's correlation, and kappa as measures of inter-examiner reliability. *Journal of Dental Research*. 1986; 65(2):128–130. [PubMed: 3455967]
64. Shriberg L, Lof G. Reliability studies in broad and narrow phonetic transcription. *Clinical Linguistics and Phonetics*. 1991; 5(3):225–279.
65. Davis B, MacNeilage P. The articulatory basis of babbling. *Journal of Speech and Hearing Research*. 1995; 38:1199–1211. [PubMed: 8747814]
66. Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*. 1960; 20(1):37–46.
67. Landis J, Koch G. The measurement of observer agreement for categorical data. *Biometrics*. 1977; 33(1):159–174. [PubMed: 843571]

68. Fleiss J. Statistical methods for rates and proportions. 2nd ed. New York: John Wiley; 1981.
69. Heaton P, Allen R, Williams K, et al. Do social and cognitive deficits curtail musical understanding? Evidence from autism and Down syndrome. *British Journal of Developmental Psychology*. 2008; 26:171–182.
70. Strand E, Stoeckel R, Baas B. Treatment of severe childhood apraxia of speech: A treatment efficacy study. *Journal of Medical Speech-Language Pathology*. 2006; 14(4):297–307.
71. Maas E, Robin D, Hula S, Austermann, et al. Principles of motor learning in treatment of motor speech disorders. *American Journal of Speech-Language Pathology*. 2008; 17:277–298. [PubMed: 18663111]
72. Iverson J, Thelen E. Hand, mouth, and brain: The dynamic emergence of speech and gesture. *Journal of Consciousness Studies*. 1999; 6:19–40.
73. Iverson J, Fagan M. Infant vocal-motor coordination: Precursor to the gesture-speech system? *Child Development*. 2004; 75(4):1053–1066. [PubMed: 15260864]
74. Collins L, Murphy S, Strecher V. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): New methods for more potent eHealth interventions. *American Journal of Preventative Medicine*. 2007; 32(5S):S112–118.
75. Kasari C, Kaiser A, Goods K, et al. Communication interventions for minimally verbal children with autism: A sequential multiple assignment randomized trial. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2014; 53(6):635–646. [PubMed: 24839882]
76. Ellis P. *The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation of Research Results*. New York: Cambridge University Press; 2010.
77. Patsopoulos N. A pragmatic view on pragmatic trials. *Dialogues in Clinical Neuroscience*. 2011; 13(2):2017–224.
78. Sox H, Goodman S. The methods of comparative effectiveness research. *Annu Rev Public Health*. 2012; 33:425–45. [PubMed: 22224891]
79. Kasari C, Paparella T, Freeman S, et al. Language outcome in autism: Randomized comparison of joint attention and play interventions. *Journal of Consulting and Clinical Psychology*. 2008; 76(1): 125–137. [PubMed: 18229990]
80. Rajadhyaksha R. Conducting feasibilities in clinical trials: An investment to ensure a good study. *Perspectives in Clinical Research*. 2010; 1(3):106–109. [PubMed: 21814631]
81. Eggen S. Generalizability of clinical trials: Why it matters for patients and public policy. *European Urology*. 2017; 71:515–516. [PubMed: 27743755]
82. Gitlin L. Introducing a new intervention: An overview of research phases and common challenges. *American Journal of Occupational Therapy*. 2013; 67:177–184. [PubMed: 23433272]

Figure 1.
Diagrammatic representation of the stages of intervention development discussed in the text.
Arrows indicate iterative refinements and bidirectional interactions between stages.

Figure 2. Comparison of the performance of the six AMMT participants from the proof-of-concept phase and the 17 AMMT participants from the replication phase. Error bars represent standard error of the mean. Statistical analysis showed no significant between-group differences.

Table 1

Participant characteristics

Study	N	Group	Sex	Age (mean, [range])	Baseline phonetic inventory ^a	Number of sessions
Proof of concept	6	AMMT	5M, 1F	6;7 [5;9–8;9]	7.9 ± 5.3	40
Replication	17	AMMT	15M, 2F	6;6 [3;5–9;8]	7.2 ± 4.3	25
Comparison	7	AMMT ^b	7M	6;1 [3;5–8;11]	7.1 ± 3.4	25
	7	SRT	5M, 2F	5;8 [3;9–8;5]	8.9 ± 5.4	25

^aPhonetic Inventory is the number of English consonants and vowels a child was able to correctly repeat (total = 31).

^bNote that these 7 AMMT participants were a subgroup of the replication group.

Table 2

Long-term comparison of AMMT and SRT (% syllables approximately correct)

	Baseline	P25	P25 + 4 weeks	P25 + 8 weeks
AMMT (<i>n</i> = 10)	27.8 ± 20.8	48.5 ± 27.7*	50.5 ± 27.8	50.2 ± 29.6
SRT (<i>n</i> = 7)	32.9 ± 20.2	37.4 ± 21.4	37.1 ± 21.1	34.3 ± 17.7

Figures are reported as mean ± standard deviation. AMMT, auditory–motor mapping training; SRT, speech repetition therapy

* $P = 0.0005$ compared with baseline. See text for details concerning the number of participants with different numbers of sessions.