Imagination exercises improve language in younger but not in older children with autism suggesting a strong critical period

Andrey Vyshedskiy[*1, 2], Edward Khokhlovich, Rita Dunn[2], Alexander Faisman, Jonah Elgart[2], Lisa Lokshina[2], Yuriy Gankin, Simone Ostrovsky[2], Lauren deTorres[2], Stephen M Edelson[3], Petr O Ilyinskii

[1]Boston University, Boston, USA

[2]ImagiRation LLC, Boston, MA, USA

[3]Autism Research Institute, San Diego, CA, USA

Correspondence and requests for materials should be addressed to vysha@bu.edu

Abstract

Based on a few cases of childhood traumatic aphasia and hemispherectomy, Lenneberg suggested a critical period for first language acquisition. Lenneberg's ideas are still debated today and have never been tested in a large group of children. Autism Spectrum Disorder (ASD) provides such an opportunity as language acquisition is a common problem in ASD. Here, we report data from a three-year-long observational trial of 8,766 children with ASD. We initiated this trial in order to test two hypotheses: 1) voluntary imagination is central to complex language and 2) there exists a strong critical period for voluntary imagination acquisition closing shortly after the age of 5. Accordingly, we developed imagination exercises (verbal and nonverbal),

organized them into an application called *Mental Imagery Therapy for Autism* (MITA) and provided this MITA app gratis to children ages 2 to 12. MITA-experienced children (N=3,540) were matched to the 'treatment-as-usual' participants (TaU, N=5,226). Both younger (2-5 years of age) and older children (5-12 YOA) in MITA and TaU groups improved their language over time, but on an annualized basis, younger MITA children improved their language 3-fold faster than TaU group. There was no difference between MITA and TaU in the older children group. These findings support Lenneberg's critical period hypothesis and indicate that acquisition of voluntary imagination is essential for the full language acquisition. Crucially, our results imply that the underlying plasticity dramatically diminishes after the age of five and therefore even greater therapeutic intervention should be targeting the very first years of a child's life.

Introduction

Full command of complex language depends on an understanding of vocabulary as well as on voluntary imagination [1] responsible for mental juxtaposition of objects into novel combinations [2]. Without voluntary imagination it is impossible to understand the difference between sentences with identical words and grammar, such as "the cat on the mat" and "the mat on the cat." Most people anthropomorphically assume innate voluntary imagination abilities in all individuals. Scientific evidence, however, points toward a more intricate story. While propensity toward voluntary imagination is innate, acquisition of voluntary imagination seems to be a function of using recursive language in early childhood [3–6]. The maturation of frontoposterior fiber tracks mediating voluntary imagination [7] depends on early childhood conversations [8,9]. In the absence of recursive conversations, children do not fine-tune these neurological connections and, as a result, do not acquire voluntary imagination to the full extent [10]. The autism community refers to

the phenomenon whereby individuals cannot combine disparate objects into a novel mental image as *stimulus overselectivity*, or *tunnel vision*, or *the lack of multi-cue responsivity* [11–13]. Failure to juxtapose mental objects results in life-long inability to understand spatial prepositions, recursion, and other complex sentences. These individuals, commonly referred to as "low-functioning," typically exhibit full-scale IQ below 70 [14,15] and usually perform below the score of 85 in non-verbal IQ tests [15]. Among individuals diagnosed with ASD, the prevalence of low-functioning ASD is 30 to 40% [16].

We hypothesized that 1) imagination exercises can improve language ability and 2) there exists a strong critical period for voluntary imagination acquisition, which ends shortly after the age of 5 [17–21]. Accordingly, we designed various developmental activities, all of which follow a systematic approach to train voluntary imagination verbally as well as outside of the verbal domain [22–25]. To make these activities dynamic and attractive to children, we organized them into an application called *Mental Imagery Therapy for Autism* (MITA). MITA verbal activities start with simple vocabulary-building exercises and progress towards exercises aimed at higher forms of language, such as noun-adjective combinations, spatial prepositions, recursion, and syntax [23]. E. g., a child can be instructed to select the {small/large} {red/ blue/green/orange} ball or to put the cup {on/under/behind/in front of} the table. All exercises are deliberately limited to as few nouns as possible since the aim is not to expand a child's one-word vocabulary, but rather to teach him/her to integrate mental objects in novel ways by utilizing voluntary imagination [23].

MITA activities outside of the verbal domain aim to provide the same voluntary imagination training visually through implicit instructions as has been described in Ref.[24]. E.g., a child can be presented with two separate images of a train and a window pattern, and a choice of complete

trains. The task is to find the correct complete train and place it into the empty square. This exercise requires not only attending to a variety of different features in both the train and its windows, but also combining two separate pieces into a single image (in other words, mentally *integrating* separate train parts into a single unified gestalt). As levels progress, the exercises increase in difficulty, requiring attention to more and more features and details. Upon attaining the most difficult levels, the child must attend to as many as eight features simultaneously. Previous results from our studies have demonstrated that children who cannot follow the explicit verbal instruction can often follow an equivalent command implicit in the visual set-up of the puzzle [23].

Voluntary imagination is an internal, subjective function that does not always manifest itself to an outside observer [2]. Unlike expanding vocabulary that can be surveyed quickly, it may take several years before an initially minimally verbal person expresses his/her voluntary imagination skills through language. As a result, studying voluntary imagination acquisition is an arduous process in which the result may not be known for several years.

We have previously described a framework for investigating targeted interventions for ASD children epidemiologically, whereby caregivers submit multiple assessments longitudinally [26]. When a single parent completes the same evaluation over multiple years, changes in the score become meaningful. Using the comprehensive 77-question Autism Treatment Evaluation Checklist (ATEC) [27] over the period of several years we have previously demonstrated significant differences between the groups of children [26]. Younger children improved more than the older children in all four ATEC subscales – Language, Sociability, Cognitive awareness, and Health. Children with milder ASD demonstrated higher improvement in the Language subscale

than children with more severe ASD. There was no difference in improvement between females vs. males in any subscale. Children from non-English-speaking countries (primarily Romance-speaking countries) improved more than children from English-speaking countries in all four subscales.

In this report we apply the same framework to study the MITA voluntary imagination intervention in children ages 2 to 12 years. The data collected over three and a half years show greater language improvement in MITA-experienced children compared to matched 'treatment as usual' controls. Crucially, this difference is observed only in children of 2 to 5 years of age implying that the underlying plasticity dramatically diminishes after the age of five and suggesting a strong critical period for voluntary imagination acquisition.

Results

We first sought to replicate our previous results [26] using the new and significantly larger databases. The analysis of groups within the TaU database confirmed the results reported in 2018. There was no difference between females vs. males in any subscale. Younger children improved more than the older children in the Language, Sociability, and Cognitive awareness subscales (Tables S1, S2). Children with milder ASD improved more than children with more severe ASD in the Language subscale (Tables S5, S6). Children from non-English-speaking countries (primarily Romance-speaking countries) improved more than children from English-speaking countries in all four subscales (Tables S9, S10). The analysis of groups within the MITA database was consistent with analysis of groups within the TaU database. There was no difference between females vs. males in any subscale. Younger children improved more than the

older children in the Language subscale (Tables S3, S4). Children with milder ASD improved more than children with more severe ASD in the Language subscale (Tables S7, S8). We did not have enough participants from non-English-speaking countries in the MITA database for statistical analysis.

Having demonstrated continuity with respect to group differences within each database, we have applied the same statistical framework to study the difference between the MITA group and the TaU group. Both younger (2-5 YOA) and older (5-12 YOA) children in the MITA and TaU groups improved their symptoms over time in all subscales. The greatest interest was the Language subscale targeted by the intervention. The average improvement in younger MITA children over three years was 8.64 points (SE=0.47, $p<0.0001$) compared to 2.88 points (SE=0.54, $p<0.0001$) in the TaU group, Figure 1, Table 1. The difference in the Language subscale in the MITA group relative to the TaU group at Visit 12 was -4.68 points (SE=0.72, $p<0.0001$); a lower score indicates greater improvement in the MITA group.

The average improvement in older MITA children over the three-year period was 3.27 points (SE=0.86, $p=0.0002$) compared to 2.41 points improvement (SE=0.28, $p<0.0001$) in the TaU group, Table 2. The difference in the MITA group relative to the TaU group at Visit 12 was not statistically significant. Thus, younger MITA children, but not older MITA children showed significantly greater improvement than the TaU group. On the annualized basis, younger MITA children improved their language three times faster than TaU children (MITA=2.88 points/year; TaU=0.96 points/year).

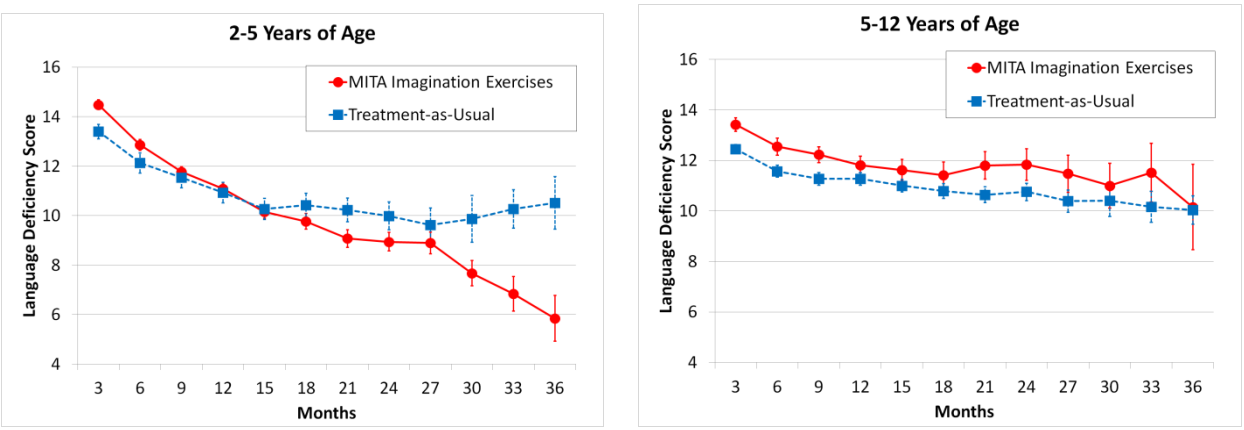

Figure 1. Language subscale score LS Means in (A) younger and (B) older children. A lower score indicates language improvement. Error bars show the 95% confidence interval.

	2-5 YOA group							
	Visit 1			Visit 12			Visit 12 - Visit 1	
	MITA	**TAU**	**MITA - TAU**	**MITA**	**TAU**	**MITA - TAU**	**MITA**	**TAU**
Subscale 1 Language	14.48 (0.1; 14.29-14.68)	13.4 (0.15; 13.11-13.7)	1.08 (0.17; <0.0001)	5.84 (0.47; 4.92-6.77)	10.52 (0.54; 9.46-11.59)	-4.68 (0.72; <0.0001)	-8.64 (0.47; <0.0001)	-2.88 (0.54; <0.0001)
Subscale 2 Sociability	14.31 (0.14; 14.04-14.58)	12.24 (0.21; 11.8-12.65)	2.07 (0.23; <0.0001)	10.55 (0.65; 9.28-11.83)	10.2 (0.75; 8.73-11.67)	0.35 (0.99; 0.7207)	-3.76 (0.65; <0.0001)	-2.05 (0.75; 0.0083)
Subscale 3 Cognitive Awareness	15.68 (0.12; 15.44-15.92)	13.89 (0.18; 13.53-14.25)	1.78 (0.2; <0.0001)	11.06 (0.58; 9.93-12.19)	9.42 (0.66; 8.12-10.72)	1.64 (0.88; 0.0628)	-4.62 (0.58; <0.0001)	-4.47 (0.66; <0.0001)
Subscale 4 Health	21.34 (0.21; 20.92-21.76)	18.91 (0.32; 18.28-19.54)	2.43 (0.36; <0.0001)	19.8 (1.01; 17.82-21.79)	15.48 (1.17; 13.19-17.77)	4.33 (1.55; 0.0052)	-1.54 (1.01; 0.1362)	-3.43 (1.17; 0.0049)

Table 1: LS Means for younger 2-5 YOA children. Data are presented as: LS Mean (SE; 95% CI). The differences between MITA and TaU and between Visit 12 and Visit 1 are presented as LS Mean (SE; P-value). A lower score indicates a lower severity of ASD symptoms.

	5-12 YOA group							
	Visit 1			Visit 12			Visit 12 - Visit 1	
	MITA	TAU	MITA - TAU	MITA	TAU	MITA - TAU	MITA	TAU
Subscale 1 Language	13.42 (0.14; 13.15-13.69)	12.45 (0.09; 12.28-12.62)	0.97 (0.16; <0.0001)	10.15 (0.86; 8.46-11.85)	10.04 (0.28; 9.48-10.6)	0.11 (0.91; 0.9013)	-3.27 (0.86; 0.0002)	-2.41 (0.28; <0.0001)
Subscale 2 Sociability	13.81 (0.19; 13.44-14.18)	12.3 (0.12; 12.07-12.53)	1.51 (0.22; <0.0001)	13.09 (1.19; 10.76-15.43)	10.26 (0.39; 9.49-11.03)	2.83 (1.25; 0.0233)	-0.72 (1.19; 0.5527)	-2.04 (0.39; <0.0001)
Subscale 3 Cognitive Awareness	14.99 (0.17; 14.66-15.32)	13.34 (0.11; 13.13-13.55)	1.65 (0.19; <0.0001)	13.72 (1.06; 11.65-15.79)	10.84 (0.35; 10.16-11.53)	2.88 (1.11; 0.0095)	-1.27 (1.06; 0.2351)	-2.5 (0.35; <0.0001)
Subscale 4 Health	21.06 (0.3; 20.48-21.64)	19.38 (0.18; 19.02-19.73)	1.68 (0.34; <0.0001)	21.9 (1.86; 18.26-25.55)	15.68 (0.61; 14.48-16.87)	6.23 (1.95; 0.0014)	0.84 (1.86; 0.6552)	-3.7 (0.61; <0.0001)

Table 2: LS Means for older 5-12 YOA children. Data are presented as: LS Mean (SE; 95% CI). The differences between MITA and TaU and between Visit 12 and Visit 1 are presented as LS Mean (SE; P-value).

On the Sociability and Cognitive awareness subscales, neither younger nor older children showed statistically significant differences in improvement between the MITA and TaU groups. On the Sociability subscale, younger MITA children improved over the three-year period by 3.76 points (SE=0.65, $p<0.0001$) compared to 2.05 points (SE=0.75, $p=0.0083$) improvement in the TaU group. The difference in the younger MITA group relative to the TaU group at Visit 12 was not statistically significant. The three-year improvement in the older MITA children was not statistically significant; older TaU children improved by 2.04 points (SE=0.39, $p<0.0001$). The difference in the older MITA group relative to the TaU group at Visit 12 was not statistically significant.

On the Cognitive awareness subscale, younger MITA children improved over the three-year period by 4.62 points (SE=0.58, $p<0.0001$) compared to 4.47 points (SE=0.66, $p<0.0001$) in the TaU group. The difference in the younger MITA group relative to the TaU group at Visit 12 was not statistically significant. The three-year improvement in the older MITA children was not statistically significant; TaU children improved by 2.50 points (SE=0.35, $p<0.0001$). The difference in the older MITA group relative to the TaU group at Visit 12 was 2.87 (SE=1.10, $p=0.01$).

On the Health subscale, younger MITA children improvement over the three-year period was not statistically significant; TaU children improved 3.43 points (SE=1.17, $p=0.0049$). The difference in the younger MITA group relative to the TaU group at Visit 12 was 4.32 (SE=1.54, $p=0.005$). The three-year improvement in the older MITA children was not statistically significant; TaU children improved by 3.70 points (SE=0.61, $p<0.0001$). The difference in the older MITA group relative to the TaU group at Visit 12 was 6.22 (SE=1.94, $p=0.001$).

Discussion

In this report, we described data from an observational trial of tablet-based imagination exercises – *Mental Imagery Therapy for Autism* or MITA [22] – that included 3,540 children with ASD who worked with MITA a median duration of 520 (IQR: 384-706) days and 5,226 treatment-as-usual (TaU) children. This is the longest-running and the largest study of a caregiver-administered early intervention tool for young children with ASD. Both younger (2-5 years of age) and older children (5-12 YOA) in MITA and TaU groups improved their language over time, but on an annualized basis, younger MITA children improved their language 3-fold faster than the TaU group, Figure 1A. There was no difference between MITA and TaU in the older children group, Figure 1B.

Greater language improvement in the younger but not older MITA children compared to TaU, supports Lenneberg's critical period hypothesis [28] independent of the exercises' exact mode of action. It is possible that imagination exercises directly trained neural networks essential for language [8]. It is also possible that MITA group caregivers were simply more motivated in administering language therapy in general. Additionally, MITA caregivers could learn language therapy techniques, such as "put the cup {on/under/behind} the table," from MITA and then extend those techniques to everyday activities multiplying the effect of exercises many-fold. The exact mechanism of MITA action cannot be identified in an observational trial. Whatever the mechanism, it worked only up to the age of five consistent with the strong critical period. Arguably, this is the most important conclusion of this study.

One of the reasons for the high number of low-functioning ASD individuals is deep misunderstanding of the critical period. It is not uncommon for parents to brush off their child's

language delay until elementary school, at which time, according to the data presented in this manuscript, it may be too late. While clinicians are usually aware of the critical period and normally recommend early intervention at the time of diagnosis, they are often reluctant to emphasize the urgent nature of the problem to the parents due to a complete lack of longitudinal studies comparing a language intervention targeting vulnerable children of various ages [29]. To the best of our knowledge, this is the first study administering the same language intervention to a large group of ASD children over three years. A better understanding of the strong critical period will result in greater effort toward language therapy in very young children and eventually in many more high-functioning productive lives.

Our second hypothesis was that voluntary imagination exercises can improve language ability. In fact, many techniques used by speech language pathologists (SLP) and Applied Behavioral Analysis (ABA) therapists happen to aim at improving voluntary imagination. SLPs commonly refer to these techniques as "combining adjectives, location/orientation, color, and size with nouns," "following directions with increasing complexity," and "building the multiple features/clauses in the sentence" [30]. In ABA jargon, these techniques are known as "visual-visual and auditory-visual conditional discrimination" [31–34], "development of multi-cue responsivity" [12], and "reduction of stimulus overselectivity" [13]. However, voluntary imagination exercises are just a small part of language therapy that primarily focuses on building up the child's vocabulary. Moreover, voluntary imagination exercises administered by therapists are mostly verbal in their nature, which can be abstruse for nonverbal and minimally-verbal children with ASD [23]. While visual reinforcers, such as 3D models and pictures, are utilized in therapy, tangible objects are always limited in their range and adaptability. Conversely, MITA exercises are disguised as

engaging games, limitless in variations, and exclusively focused on training voluntary imagination.

The hypothesis that voluntary imagination exercises can improve language ability would have been rejected if data presented herein had shown no difference between MITA and TaU groups. This has not been the case since our data showed significantly greater improvement in language in younger children which have used MITA. This observation can be explained by the direct or indirect effect of exercises or by selection-bias of MITA caregivers, who could simply be more motivated in their approach to therapy. In fact, MITA caregivers had to actively search for autism apps, download the app and then administer the therapy over six months or longer. However, TaU caregivers have also had to demonstrate similar motivation by actively searching for online ATEC evaluation and voluntarily completing at least three evaluations over 12 months or longer. Furthermore, if greater caregiver motivation is the explanation for better MITA outcome, one would expect to see an improved outcome in all areas of child development. However, MITA group showed better outcome only in the Language subscale, but not in the Sociability, Cognitive awareness, or Health subscales. Moreover, the greatest difference in language is only expressed two years after beginning of MITA exercises, see Figure 1A. Improvements in vocabulary can usually be detected much faster. The reported long delay is consistent with the initial improvement of an internal component, such as the subjective function of voluntary imagination, that does not readily manifest itself to an outside observer [2]. The Language subscale of ATEC primarily assesses expressive language that depends on voluntary imagination, but can be influenced by many other factors as well. It may take several years before an initially nonverbal or minimally verbal child expresses his/her voluntary imagination

skills through language. We conclude that the direct or indirect effect of imagination exercises on language cannot be excluded and in fact could be the most parsimonious explanation for the observed results.

Limitations

The observational design of this study cannot definitively prove causality since unknown confounders may influence the study results. The golden standard of testing a novel clinical intervention is randomized controlled trial (RCT). Prior to conducting the MITA study, we have run the proposal for a therapist-administered RCT of voluntary imagination intervention through many potential funders and collaborators. The proposal has failed to find any traction. We have also considered a caregiver-administered RCT, but decided against it due to high attrition rate. The only published RCT of caregiver-administered tablet-based therapy for young children with ASD reported an overwhelming drop just after 3 months despite every biweekly telephone calls to encourage app use [35]: during the first 3-month period, participants exercised for a total median time of 1,593 minutes (just under the recommended target of 20 min/day or 1,800 min/3-month period); during the second 3-month period, participants exercised for a total median time of 23 minutes (98.6% drop in app use). In effect, most participants did not receive any intervention after the first 3-month period and therefore were lost for the RCT [35]. As the minimal length of an voluntary imagination intervention RCT is likely to exceed two years [36,37], participant dropout becomes the major issue. This high attrition rate introduces multiple selection biases that degrade RCT ability to demonstrate causality and essentially makes it no better than an observational trial. In this study, we used propensity score analysis [38] to identify comparable individuals based on age and all four evaluation subscales. For each participant in the MITA group, a match was

found by choosing the control observation with the closest propensity score. Propensity score matching does not completely address selection bias, but it significantly improves group similarity.

Another disadvantage of low-cost geographically diverse observational trials is their reliance on parent-reported outcome measures. There is an understanding in the psychological community that parents cannot be trusted with an evaluation of their own children. In fact, parents often yield to wishful thinking and overestimate their children's abilities on a single assessment [39]. However, the pattern of changes can be generated by measuring the score dynamics over multiple assessments. When a single parent completes the same evaluation every three months over multiple years, changes in the score become meaningful. In this trial we used a comprehensive 77-question ATEC evaluation validated in multiple clinical trials [26,35,40–45], assessing participants on four diverse scales of development (Language, Sociability, Cognitive awareness, and Health) and administered at regular 3-month intervals. If there was a selection bias in parents' responses in this trial, it would be expected to artificially inflate improvement in the TaU group. Unlike MITA parents, evaluation reminders were not sent to TaU parents. All TaU evaluations have been completed spontaneously and therefore may have experienced a selection bias among parents for whom the evaluations reported greater improvement. This selection bias would not be expected in the MITA group, as all evaluations were scheduled and mandatory.

Conclusions

Five major conclusions follow up from this study. First, the plasticity essential for acquisition of

first language may reduce significantly after the age of five years consistent with Lenneberg's critical period hypothesis [28]. Second, voluntary imagination is an essential component of full language and imagination exercises may be an indispensable component of language therapy. Third, the minimal necessary duration of a clinical trial investigating the effect of imagination exercises could exceed two years. Fourth, some caregivers are capable of administering tablet-based exercises to their children consistently over many years. Fifth, parent-administered parent-reported multiyear observational trials can be an attractive low-cost model for studying novel language, behavioral, and dietary interventions. The significant improvement of language observed in the current trial brings hope to many families and inspires us to continue developing imagination exercises and translate MITA to multiple language. The major strength of this study is the large number of long-term participants. The most obvious limitation of the study is that this study observational design cannot definitively prove causality since not all confounders can be adjusted appropriately.

Methods

MITA exercises

MITA includes both verbal and nonverbal exercises aiming to develop voluntary imagination ability [3]. MITA verbal activities use higher forms of language, such as noun-adjective combinations, spatial prepositions, recursion, and syntax [4] to train voluntary imagination: e. g., a child can be instructed to put the *large red dog behind the orange chair*, Figure 2A; or *identify the wet animal* after *the lion was showered by the monkey*; or *take animals home* following an explanation that *the lion lives above the monkey and under the cow*, Figure 2B. In every activity

a child listens to a short story, then works within an immersive interface to generate an answer; correct answers are rewarded. To avoid routinization, all instructions are generated dynamically from individual words. Collectively, verbal activities have over 10 million different instructions, so most instructions will never be heard twice by a child.

Figure 2. Examples of MITA verbal exercises. (A) A child is instructed to put the *large red dog behind the orange chair*. (B) A child is instructed: *Imagine. The lion lives above the monkey and under the cow. Take animals home.* Note that animals cannot be dragged to their apartments during instructions, encouraging a child to imagine animals' correct positions in the mind.

MITA nonverbal activities aim to provide the same voluntary imagination training visually through implicit instructions [5]. E.g., a child can be presented with two separate images of a train and a window pattern, and a choice of complete trains. The task is to find the correct complete train. The child is encouraged to avoid trial-and-error and integrate separate train parts mentally,

thus training voluntary imagination, Figure 3A. Different games use various tasks and visual patterns to keep a child engaged, Figure 3B. Most puzzles are assembled dynamically from multiple pieces so that they never repeat themselves. Collectively, MITA activities are designed to last for approximately 10 years.

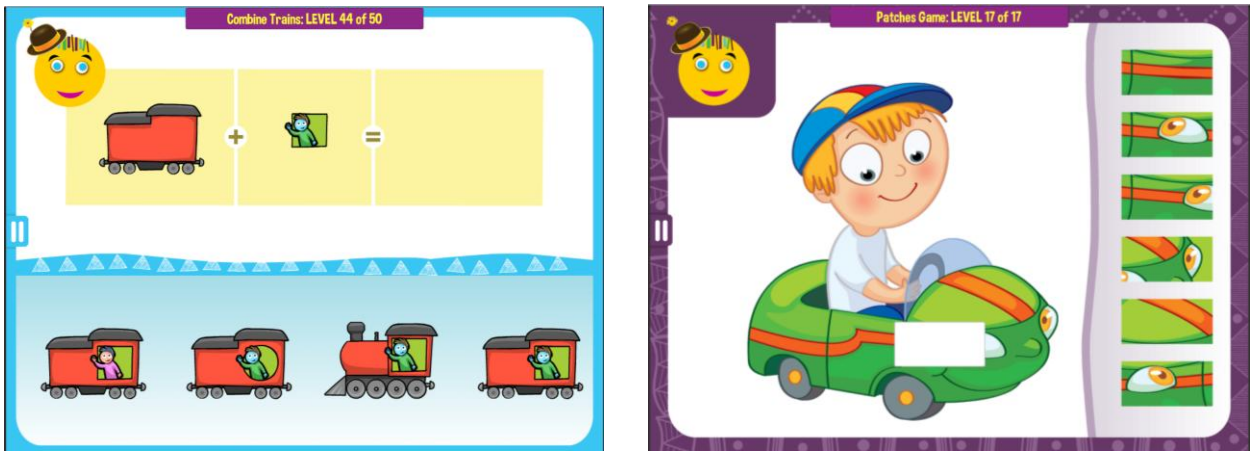

Figure 3. Examples of MITA nonverbal exercises. (A) Implicit instruction: *Find the correct train.* (B) Implicit instruction: *Find the correct patch.*

MITA group

The MITA app was made available gratis at all major app stores in February 2016. Once the app was downloaded, the caregiver was asked to register and to provide demographic details, including the child's diagnosis and age. Caregivers consented to anonymized data analysis and completed Autism Treatment Evaluation Checklist (ATEC) [27]. The first evaluation was administered approximately one month after the first use of MITA and once 100 puzzles had been completed. The subsequent evaluations were administered at three-month intervals. Parents

were asked to complete evaluations independently of a child's actual use of MITA.

From this pool of potential study participants, we selected participants based on the following criteria:

1) Consistency: Participants must have filled out at least three ATEC evaluations and the interval between the first and the last evaluation was six months or longer.

2) Diagnosis: The subject must have self-reported their diagnosis as ASD.

3) Maximum age: Participants older than twelve years of age were excluded from this study.

4) Minimum age: Participants who completed their first evaluation before the age of two years were excluded from this study.

5) Minimal ATEC severity: Participants with initial ATEC scores of less than 20 were excluded.

6) Language: Participants who indicated their primary language was not English were excluded from the study.

After excluding participants that did not meet these criteria, there were 3,540 total participants, Table 3.

Control group

Independently from MITA, ATEC responses were collected by the Autism Institute from participants voluntarily completing online ATEC evaluations from 2013 to 2019. Little is known about their treatment, but it is unlikely that many of them used MITA. Accordingly, these participants served as a 'treatment as usual' control. Participant selection was described in detail

in Ref. [26]. In short, participants were selected based on the following criteria:

1) Completeness: Participants who did not provide a date of birth (DOB) were excluded. As participants' DOB were utilized to determine age, the availability of DOB was necessary.

2) Consistency: Participants had to have completed at least three questionnaires and the interval between the first and the last evaluation was one year or longer.

3) Maximum age: Participants older than twelve years of age were excluded from this study.

As diagnosis was not part of the ATEC questionnaire, some neurotypical participants could be present in the database. To limit the contribution from neurotypical children, we excluded participants that may have represented the neurotypical population by using the *Minimum age* and *the Minimal ATEC severity* criteria.

4) Minimum age: Participants who completed their first evaluation before the age of 2 were excluded from this study, as the diagnosing of ASD in this age group is uncertain and the parents of some of these young cases may have completed the ATEC because they wanted to check whether their normal child had signs of autism.

5) Minimal ATEC severity: Participants with initial ATEC scores of less than 20 were excluded.

6) **Language**: Participants who indicated their primary language was not English were excluded from the study.

After excluding participants that did not meet these criteria, there were 5,226 total participants.

	Age	Participants in each age group (total)	Participants in each age group (%)	Age at baseline (mean ± SD)	Initial ATEC score (mean ± SD)
MITA	2-5 YOA	2614	73.84%	3.53 ± 0.76	75.75 ± 23.14
	5-12 YOA	926	26.16%	7.10 ± 1.84	75.28 ± 21.86
	Total	**3540**	**100.00%**	**4.46 ± 1.94**	**75.63 ± 22.81**
TaU	2-5 YOA	2558	48.95%	3.57 ± 0.79	63.31 ± 23.29
	5-12 YOA	2668	51.05%	7.39 ± 1.90	58.16 ± 23.54
	Total	**5226**	**100.00%**	**5.52 ± 2.41**	**60.68 ± 23.56**

Table 3. Characteristics and baseline measures for all age groups. A lower ATEC score indicates a lower severity of ASD symptoms.

Outcome measures

A caregiver-completed Autism Treatment Evaluation Checklist (ATEC) [27] was used to track the efficacy of a treatment. ATEC is comprised of four subscales: 1) Speech/Language/Communication, 2) Sociability, 3) Sensory/Cognitive Awareness, and 4) Physical/Health/Behavior. The first subscale, Speech/Language/Communication, contains 14 items and its score ranges from 0 to 28 points. The Sociability subscale contains 20 items within a score range from 0 to 40 points. The third subscale, Sensory/Cognitive awareness, has 18 items and scores range from 0 to 36 points. Finally, the Health/Physical/Behavior subscale contains 25 items and scores range from 0 to 75 points. The scores from each subscale are combined in order to calculate a Total Score, which ranges from 0 to 179 points. A lower score indicates lower severity of ASD symptoms and a higher score correlates with more severe symptoms of ASD.

Statistical analysis

The framework for evaluation of ATEC score changes over time was explained in detail in Ref.

[26]. In short, the concept of a "Visit" was developed by dividing the three-year-long observation interval into 3-month periods. All evaluations were mapped into 3-month-long bins with the first evaluation placed in the first bin. When more than one evaluation was completed within a bin, their results were averaged to calculate a single number representing this 3-month interval. It was then hypothesized that there was a three-way interaction between an age group, Visit, and treatment. Statistically, this hypothesis was modeled by applying the Linear Model with repeated measures, where a three-way interaction term was introduced to test the hypothesis. Least squares means (LS Means) and LS Means differences were calculated for all ATEC subscales (Language, Sociability, Cognitive awareness, and Health) at all visits. Participants in the MITA group were matched to those in TaU group using propensity score analysis [38] based on age and all four ATEC subscales at baseline.

Informed Consent

Caregivers have consented to anonymized data analysis and publication of the results.

Clinical Trial Registration

The observational clinical trial, ClinicalTrials.gov Identifier: NCT02708290, was registered on March 15, 2016.

Compliance with Ethical Standards

Using the Department of Health and Human Services regulations found at 45 CFR 46.101(b)(1), it was determined that this research project is exempt from IRB oversight.

Data Availability

De-identified raw data from this manuscript are available from the corresponding author upon reasonable request.

Code availability statement

Code is available from the corresponding author upon reasonable request.

Acknowledgments

We wish to thank Danielle Abate for assistance with manuscript preparation and Yulia Dumov for design of children activities.

Author contributions

AV, EK, and POI designed the study. RD, AF, JE, LL, YG and AV developed the MITA app. SME acquired the treatment as usual data. EK, SO, LdT, and AV analyzed the data. AV, EK, and POI wrote the paper.

Competing Interests

This study was self-funded. AV, EK, POI, RD, AF, JE, LL, and YG are partners in ImagiRation limited liability partnership, the developer of MITA.

References

1. Pearson, J. The human imagination: the cognitive neuroscience of visual mental imagery. *Nat. Rev. Neurosci.* 1–11 (2019).

2. Vyshedskiy, A. Neuroscience of imagination and implications for hominin evolution. *J. Curr. Neurobiol.* (2019). doi:10.31234/osf.io/skxwc

3. Martin, A. J. Does age of language acquisition affect the relation between American sign language and mental rotation? (UNIVERSITY OF MINNESOTA, 2009).

4. Martin, A., Senghas, A. & Pyers, J. Age of acquisition effects on mental rotation: evidence from Nicaraguan sign language. in *BUCLD 37: proceedings of the 37th Boston university conference on language development* 241–250 (2013).

5. Pyers, J. E., Shusterman, A., Senghas, A., Spelke, E. S. & Emmorey, K. Evidence from an emerging sign language reveals that language supports spatial cognition. *Proc. Natl. Acad. Sci.* **107**, 12116–12120 (2010).

6. Bick, J. *et al.* Effect of early institutionalization and foster care on long-term white matter development: a randomized clinical trial. *JAMA Pediatr.* **169**, 211–219 (2015).

7. Skeide, M. A., Brauer, J. & Friederici, A. D. Brain functional and structural predictors of language performance. *Cereb. Cortex* **26**, 2127–2139 (2015).

8. Romeo, R. R. *et al.* Language exposure relates to structural neural connectivity in childhood. *J. Neurosci.* **38**, 7870–7877 (2018).

9. Cheng, Q., Roth, A., Halgren, E. & Mayberry, R. I. Effects of early language deprivation on

brain connectivity: Language pathways in deaf native and late first-language learners of American Sign Language. *Front. Hum. Neurosci.* **13**, 320 (2019).

10. Vyshedskiy, A., Mahapatra, S. & Dunn, R. Linguistically deprived children: meta-analysis of published research underlines the importance of early syntactic language use for normal brain development. *Res. Ideas Outcomes* (2017). doi:10.3897/rio.3.e20696

11. Schreibman, L. Diagnostic features of autism. *J. Child Neurol.* **3**, S57–S64 (1988).

12. Lovaas, O. I., Koegel, R. L. & Schreibman, L. Stimulus overselectivity in autism: A review of research. *Psychol. Bull.* **86**, 1236–1254 (1979).

13. Ploog, B. O. Stimulus overselectivity four decades later: A review of the literature and its implications for current research in autism spectrum disorder. *J. Autism Dev. Disord.* **40**, 1332–1349 (2010).

14. Beglinger, L. J. & Smith, T. H. A review of subtyping in autism and proposed dimensional classification model. *J. Autism Dev. Disord.* **31**, 411–422 (2001).

15. Boucher, J., Mayes, A. & Bigham, S. Memory, language and intellectual ability in low-functioning autism. *J Boucher DM Bowler Eds Mem. Autism Camb. CUP* (2008).

16. Fombonne, E. Epidemiological surveys of autism and other pervasive developmental disorders: an update. *J. Autism Dev. Disord.* **33**, 365–382 (2003).

17. Boatman, D. *et al.* Language recovery after left hemispherectomy in children with late-onset seizures. *Ann. Neurol.* **46**, 579–586 (1999).

18. Basser, L. S. Hemiplegia of early onset and the faculty of speech with special reference to

the effects of hemispherectomy. *Brain* **85**, 427–460 (1962).

19. Lenneberg, E. H. The biological foundations of language. *Hosp. Pract.* **2**, 59–67 (1967).

20. Krashen, S. & Harshman, R. Lateralization and the critical period. *J. Acoust. Soc. Am.* **52**, 174–174 (1972).

21. Pulsifer, M. B. *et al.* The cognitive outcome of hemispherectomy in 71 children. *Epilepsia* **45**, 243–254 (2004).

22. Vyshedskiy, A. & Dunn, R. Mental Imagery Therapy for Autism (MITA)-An Early Intervention Computerized Brain Training Program for Children with ASD. *Autism Open Access* **5**, 2 (2015).

23. Dunn, R. *et al.* Comparison of performance on verbal and nonverbal multiple-cue responding tasks in children with ASD. *Autism Open Access* **7**, 218 (2017).

24. Dunn, R. *et al.* Children With Autism Appear To Benefit From Parent-Administered Computerized Cognitive And Language Exercises Independent Of the Child's Age Or Autism Severity. *Autism Open Access* **7**, (2017).

25. Dunn, R. *et al.* Tablet-Based Cognitive Exercises as an Early Parent-Administered Intervention Tool for Toddlers with Autism - Evidence from a Field Study. *Clin. Psychiatry* **3**, (2017).

26. Mahapatra, S. *et al.* Longitudinal Epidemiological Study of Autism Subgroups Using Autism Treatment Evaluation Checklist (ATEC) Score. *Autism Dev. Disord.* **1**, (2018).

27. Rimland, B. & Edelson, S. Autism Research Institute. *Autism Treat. Eval. Checkl. ATEC*

(1999).

28. Lenneberg, E. H. The biological foundations of language. *Hosp. Pract.* **2**, 59–67 (1967).

29. WEBB, S. J., JONES, E. J. H., KELLY, J. & DAWSON, G. The motivation for very early intervention for infants at high risk for autism spectrum disorders. *Int. J. Speech Lang. Pathol.* **16**, 36–42 (2014).

30. American Speech-Language-HearingAssociation. Scope of practice in speech-language pathology. (2016).

31. Axe, J. B. Conditional discrimination in the intraverbal relation: A review and recommendations for future research. *Anal. Verbal Behav.* **24**, 159–174 (2008).

32. Michael, J., Palmer, D. C. & Sundberg, M. L. The multiple control of verbal behavior. *Anal. Verbal Behav.* **27**, 3–22 (2011).

33. Eikeseth, S. & Smith, D. P. An analysis of verbal stimulus control in intraverbal behavior: implications for practice and applied research. *Anal. Verbal Behav.* **29**, 125–135 (2013).

34. Lowenkron, B. Joint control and the selection of stimuli from their description. *Anal. Verbal Behav.* **22**, 129–151 (2006).

35. Whitehouse, A. J. *et al.* A randomised controlled trial of an iPad-based application to complement early behavioural intervention in Autism Spectrum Disorder. *J. Child Psychol. Psychiatry* (2017).

36. Dawson, G. *et al.* Randomized, controlled trial of an intervention for toddlers with autism: the Early Start Denver Model. *Pediatrics* **125**, e17–e23 (2010).

37. Smith, T., Groen, A. D. & Wynn, J. W. Randomized trial of intensive early intervention for children with pervasive developmental disorder. *Am. J. Ment. Retard.* **105**, 269–285 (2000).

38. Schneider, B., Carnoy, M., Kilpatrick, J., Schmidt, W. H. & Shavelson, R. J. *Estimating causal effects using experimental and observational design*. (American Educational & Reseach Association, 2007).

39. Scattone, D., Raggio, D. J. & May, W. Comparison of the vineland adaptive behavior scales, and the bayley scales of infant and toddler development. *Psychol. Rep.* **109**, 626–634 (2011).

40. Al Backer, N. B. Correlation between Autism Treatment Evaluation Checklist (ATEC) and Childhood Autism Rating Scale (CARS) in the evaluation of autism spectrum disorder. *Sudan. J. Paediatr.* **16**, 17 (2016).

41. Geier, D. A., Kern, J. K. & Geier, M. R. A comparison of the Autism Treatment Evaluation Checklist (ATEC) and the Childhood Autism Rating Scale (CARS) for the quantitative evaluation of autism. *J. Ment. Health Res. Intellect. Disabil.* **6**, 255–267 (2013).

42. Jarusiewicz, B. Efficacy of Neurofeedback for Children in the Autistic Spectrum: A Pilot Study. *J. Neurother.* **6**, 39–49 (2002).

43. Magiati, I., Moss, J., Yates, R., Charman, T. & Howlin, P. Is the Autism Treatment Evaluation Checklist a useful tool for monitoring progress in children with autism spectrum disorders? *J. Intellect. Disabil. Res.* **55**, 302–312 (2011).

44. Charman, T., Howlin, P., Berry, B. & Prince, E. Measuring developmental progress of children with autism spectrum disorder on school entry using parent report. *Autism* **8**, 89–100 (2004).

45. Klaveness, J., Bigam, J. & Reichelt, K. L. The varied rate of response to dietary intervention in autistic children. *Open J. Psychiatry* **3**, 56 (2013).