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

Abstract

Imagination exercises administered by caregivers were investigated in a three-year-long observational trial of 3,540 children with autism aged 2-12 years. Tablet-based verbal and nonverbal exercises modeled on language therapy and emphasizing mental-juxtaposition-of-objects were organized into an application called *Mental Imagery Therapy for Autism* (MITA). MITA-exposed children were matched to the 'Treatment-as-Usual' participants (TaU, N=5,222) by age, gender, language, sociability, cognitive awareness, health, and ASD severity at baseline. Both younger (2-5 years-of-age) and older children (5-12 YOA) in MITA and TaU groups improved their symptoms over time, but younger MITA-exposed children showed 2.3-fold improvement in language score at the end of the trial vs. TaU group. There was no difference between MITA and TaU in the older children group, supporting Lenneberg's critical period hypothesis.

*Corresponding author: Andrey Vyshedskiy, Ph.D., Boston University, Boston, USA, Tel: +1 (617) 433-7724; E-mail: vysha@bu.edu

Keywords: autism, ASD, language delay, language therapy, imagination, theory of mind

Introduction

Full command of complex language depends on understanding of vocabulary as well as on voluntary imagination (Pearson, 2019) responsible for mental juxtaposition of objects into novel combinations (Vyshedskiy, 2019). 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, suggests a more intricate story. While propensity toward voluntary imagination is innate, acquisition of voluntary imagination seems to be the function of using recursive language in early childhood (Bick et al., 2015; A. J. Martin, 2009; A. Martin, Senghas, & Pyers, 2013; Pyers, Shusterman, Senghas, Spelke, & Emmorey, 2010). The myelination of frontoposterior fiber tracts mediating voluntary imagination (Skeide, Brauer, & Friederici, 2015) depends on early childhood conversations (Cheng, Roth, Halgren, & Mayberry, 2019; Romeo et al., 2018). 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 (Vyshedskiy, Mahapatra, & Dunn, 2017). 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* (Dube et al., 2016; Lovaas, Koegel, & Schreibman, 1979; Ploog, 2010; Schreibman, 1988). Failure to juxtapose mental objects results in life-long inability to understand spatial prepositions, recursion, and other complex sentences. Among individuals diagnosed with ASD, the prevalence of individuals exhibiting this problem is 30 to 40% (Fombonne, 2003).

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. 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 (R. Dunn, Elgart, Lokshina, Faisman, Khokhlovich, et al., 2017b, 2017a; R. Dunn, Elgart, Lokshina, Faisman, Waslick, et al., 2017; Vyshedskiy & Dunn, 2015a). To make these activities dynamic and attractive to children, we organized them into an application called *Mental Imagery Therapy for Autism* (MITA). The MITA app release in 2016 was an instant success with nearly 100,000 downloads each year.

MITA verbal activities start with simple vocabulary-building exercises and progress toward exercises aimed at higher forms of language, such as noun-adjective combinations, spatial prepositions, recursion, and syntax (R. Dunn, Elgart, Lokshina, Faisman, Khokhlovich, et al., 2017b). 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 (R. Dunn, Elgart, Lokshina, Faisman, Khokhlovich, et al., 2017b).

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. (R. Dunn, Elgart, Lokshina, Faisman, Khokhlovich, et al., 2017a). 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 (R. Dunn, Elgart, Lokshina, Faisman, Khokhlovich, et al., 2017b).

Voluntary imagination is an internal, subjective function that does not immediately manifest itself to parents and caregivers (Vyshedskiy et al., 2020). Unlike expanding vocabulary that can be surveyed quickly, it may take several years before an initially minimally verbal child expresses his/her voluntary imagination skills through language. As a result, randomized controlled trial (RCT) of voluntary imagination intervention is an arduous proposition that may take more than two years and will necessitate a large number of participants to balance inevitable dropout. We were not able to generate support for an RCT and therefore resolved to conduct a simpler observational trial.

We have previously described a framework for investigating targeted interventions for ASD children epidemiologically, whereby caregivers submit multiple assessments longitudinally (Mahapatra, Khokhlovich, et al., 2018). When a single parent completes the same evaluation over multiple years, changes in the score become a meaningful measure of child's progress

despite a possible bias in the absolute score. Using the comprehensive 77-question Autism Treatment Evaluation Checklist (ATEC) (Rimland & Edelson, 1999) over the period of several years we have demonstrated significant differences in outcomes along several parameters (Mahapatra, Khokhlovich, et al., 2018). 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. Children from non-English-speaking countries (primarily Romance-speaking countries) improved more than children from English-speaking countries in all four subscales. No difference between females vs. males was registered in all cohorts studied.

In this report we apply the same framework to study MITA voluntary imagination intervention in children aged 2 to 12 years. The data collected over three and a half years shows greater language improvement in MITA-exposed 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 component of language.

Methods

MITA exercises

MITA includes both verbal and nonverbal exercises aiming to develop voluntary imagination ability (Pearson, 2019). The fidelity, validity and reliability of the MITA was discussed in detail in Refs. (R. Dunn, Elgart, Lokshina, Faisman, Khokhlovich, et al., 2017b, 2017a; R. Dunn,

Elgart, Lokshina, Faisman, Waslick, et al., 2017; Vyshedskiy & Dunn, 2015a). MITA verbal activities use higher forms of language, such as noun-adjective combinations, spatial prepositions, recursion, and syntax (R. Dunn, Elgart, Lokshina, Faisman, Khokhlovich, et al., 2017b) to train voluntary imagination: e. g., a child can be instructed to put the *large red dog behind the orange chair*, Figure 1A; 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 1B. In every activity a child listens to a short story, then works within an immersive interface to generate an answer; correct answers are rewarded with pre-recorded encouragement and flying stars. To avoid routinization, all instructions are generated dynamically from individual words. Collectively, verbal activities have over 10 million different instructions, therefore a child will almost never hear the same instruction once again.

Figure 1. 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 (R. Dunn, Elgart, Lokshina, Faisman, Khokhlovich, et al., 2017a). 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 2A. Different games use various tasks and visual patterns to keep the child engaged, Figure 2B. Most puzzles are assembled dynamically from multiple pieces in such a way that they never repeat themselves.

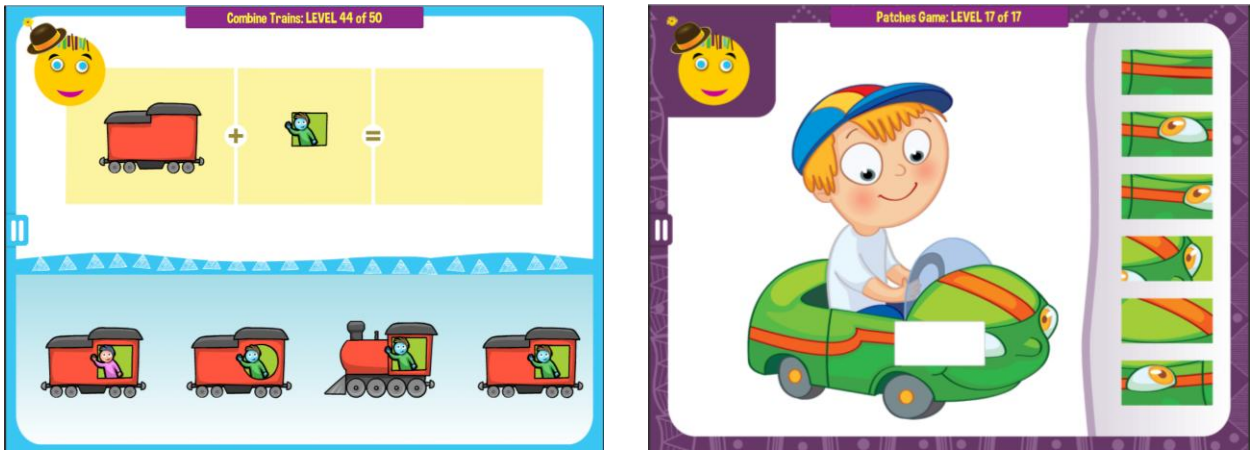

Figure 2. Examples of MITA nonverbal exercises. (A) Implicit instruction: *Find the correct train.* (B) Implicit instruction: *Find the correct patch.*

MITA also includes a number of hybrid activities that start children on easier nonverbal exercises and then gradually increase in difficulty, first to a combination of a verbal instruction

and a visual clue and later to a verbal instruction alone, Figure 3. Collectively, MITA activities are designed to last for approximately 10 years.

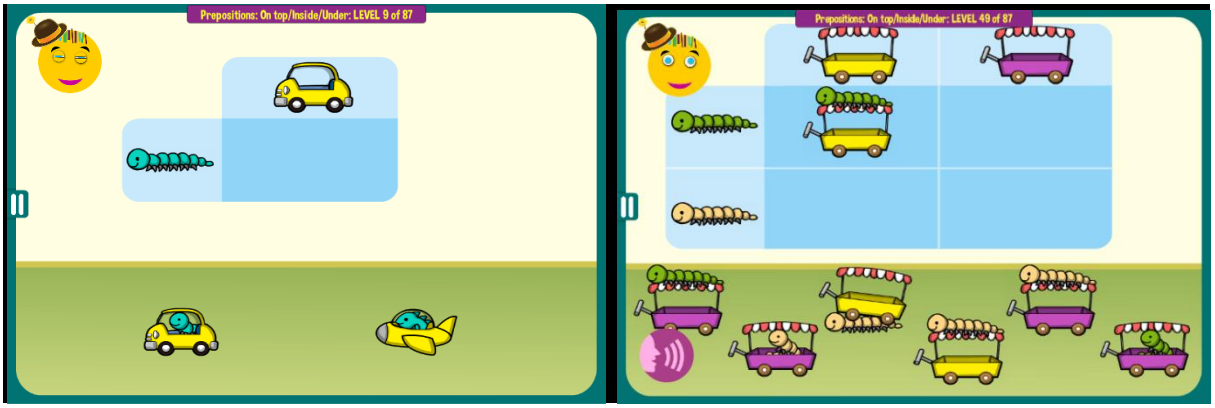

Figure 3. Examples of a MITA game that teach spatial prepositions *above* and *under*. (A) The game starts with the implicit instruction to combine an animal and a vehicle. (B) At more difficult levels a child must notice the correct positioning of the animal (*above* or *under*), which is announced verbally and also indicated by a visual clue in the top left corner of the matrix. At the most difficult levels (not shown) the visual clue is hidden and a child must rely on the verbal instruction alone.

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) (Rimland & Edelson, 1999). 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

approximately three-month intervals. To enforce regular evaluations, MITA app became unusable at the end of each three-month interval and parents needed to complete an evaluation to regain its functionality.

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 subjects must have self-reported their diagnosis as ASD. The ASD diagnosis was not verified directly, as we cannot ask participants to submit documentation. However, ATEC scores support ASD diagnosis. Average initial ATEC total score in the MITA group was 75.63 ± 22.81, Table 1, which corresponds to medium-to-severe ASD as delineated in Ref. (Mahapatra, Vyshedsky, et al., 2018) and Table 2.

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 to further limit the contribution from neurotypical children.

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 1.

Control group

Independently from MITA, ATEC responses were collected by the Autism Institute from participants voluntarily completing online ATEC evaluations from 2013 to 2019. These participants were cross-checked against the MITA group to ensure that none of them used MITA. Accordingly, these participants served as a 'treatment-as-usual' control. Participant selection was described in detail in Ref. (Mahapatra, Khokhlovich, et al., 2018). 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 subjects 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 5,222 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	48.99%	3.57 ± 0.79	63.31 ± 23.29
	5-12 YOA	2664	51.02%	7.39 ± 1.90	58.17 ± 23.55
	Total	**5222**	**100.00%**	**5.52 ± 2.41**	**60.69 ± 23.56**

Table 1. 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) (Rimland & Edelson, 1999) was used to track the efficacy of the treatment. The complete ATEC questionnaire can be accessed freely at www.autism.org. 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 indicates more severe symptoms of ASD. ATEC is not a diagnostic checklist. It was designed to evaluate treatment effectiveness (Rimland & Edelson, 1999) and ASD severity can be related to ATEC total score and age only approximately. Table 2 lists approximate ATEC total score as related to ASD severity and age as described elsewhere (Mahapatra, Khokhlovich, et al., 2018).

Table 2. Approximate relationship between ATEC total score, age, and ASD severity as described elsewhere (Mahapatra, Khokhlovich, et al., 2018)

	Age										
Severity	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Mild	<82	<65	<52	<43	<36	<31	<28	<25	<23	<21	<20
Moderate	82-130	65-103	52-83	43-69	36-58	31-50	28-44	25-39	23-36	21-34	20-32
Severe	130-179	103-179	83-179	69-179	58-179	50-179	44-179	39-179	36-179	34-179	32-179

ATEC was selected because it is one of the few measures validated to evaluate treatment effectiveness. In contrast, another popular ASD assessment tool, ADOS, (Lord et al., 2000) has been only validated as a diagnostic tool. Various studies confirmed validity and reliability of ATEC (Al Backer, 2016; Geier, Kern, & Geier, 2013; Jarusiewicz, 2002) and several trials confirmed ATEC's ability to longitudinally measure changes in participant performance (Charman, Howlin, Berry, & Prince, 2004; Klaveness, Bigam, & Reichelt, 2013; Magiati, Moss, Yates, Charman, & Howlin, 2011; Mahapatra, Khokhlovich, et al., 2018). Whitehouse et al. used ATEC as a primary outcome measure for a randomized controlled trial of their iPad-based intervention for ASD named TOBY and noted ATEC's "internal consistency and adequate predictive validity" (Whitehouse et al., 2017). These studies support the viability of ATEC as a tool for longitudinal tracking of symptoms and assessing changes in ASD severity.

ATEC Language subscale

ATEC language subscale consists of 14 items. These items start by assessing the simplest linguistic abilities, such as 1) *Knows own name*, 2) *Responds to 'No' or 'Stop'*, 3) *Can follow some commands*, 4) *Can use one word at a time (No!, Eat, Water, etc.)*, 5) *Can use 2 words at a time (Don't want, Go home)*, 6) *Can use 3 words at a time (Want more milk)* and progress to interrogate complex language abilities, such as 7) *Knows 10 or more words* , 8) *Can use sentences with 4 or more words*, 9) *Explains what he/she wants* , 10) *Asks meaningful questions*, 11) *Speech tends to be meaningful/relevant* (item 11), *Often uses several successive sentences*, 12) *Carries on fairly good conversation*, and 14) *Has normal ability to communicate for his/her age*. With the exception of the first three items, all the Language subscale items depend on expressive language. This is expected as ATEC is a parents' questionnaire and has to rely on answers that are familiar to parents and cannot involve any special tests.

Statistical analysis

The framework for evaluation of ATEC score changes over time was explained in detail earlier (Mahapatra, Khokhlovich, et al., 2018). 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 Mixed Effect Model with Repeated Measures (MMRM), where a three-

way interaction term was introduced to test the hypothesis. The model (Endpoint ~ Baseline + Gender + Severity + Age_Group * Treatment * Visit) was fit using R Bioconductor library of statistical packages, in particular “nlme” package. The subscale score at baseline, gender, and severity were used as covariates. Conceptually, the model fits a plane into n-dimensional space. This plane takes into account a complex variability structure across multiple visits, including baseline differences. Once such plane is fit, the model calculates Least Squares Means (LS Means) for each subscale, treatment group, and age group at each visit. The model also calculates LS Mean differences between the treatment and control groups at each visit. Participants in the MITA group were matched to those in the TaU group using propensity score analysis (Schneider, Carnoy, Kilpatrick, Schmidt, & Shavelson, 2007) based on age, gender, ASD severity, 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. The results have been posted on ClinicalTrials.gov on December 20, 2019.

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.

Results

We first sought to replicate our earlier results (Mahapatra, Khokhlovich, et al., 2018) 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). 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 8.40 points (SE=0.42, $p<0.0001$) compared to 3.66 points (SE=0.41, $p<0.0001$) in the TaU group, Figure 4A, Table 3. The difference in the Language subscale in the MITA group relative to the TaU group at Visit 12 was -3.94 points (SE=0.59, $p<0.0001$); a lower score indicates greater improvement in the MITA group.

The average improvement in older MITA children over the three-year period was 2.18 points (SE=0.77, $p=0.0046$) compared to 2.45 points improvement (SE=0.28, $p<0.0001$) in the TaU group, Figure 4B, Table 4. 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 2.3-times faster than TaU children (MITA=2.8 points/year; TaU=1.2 points/year).

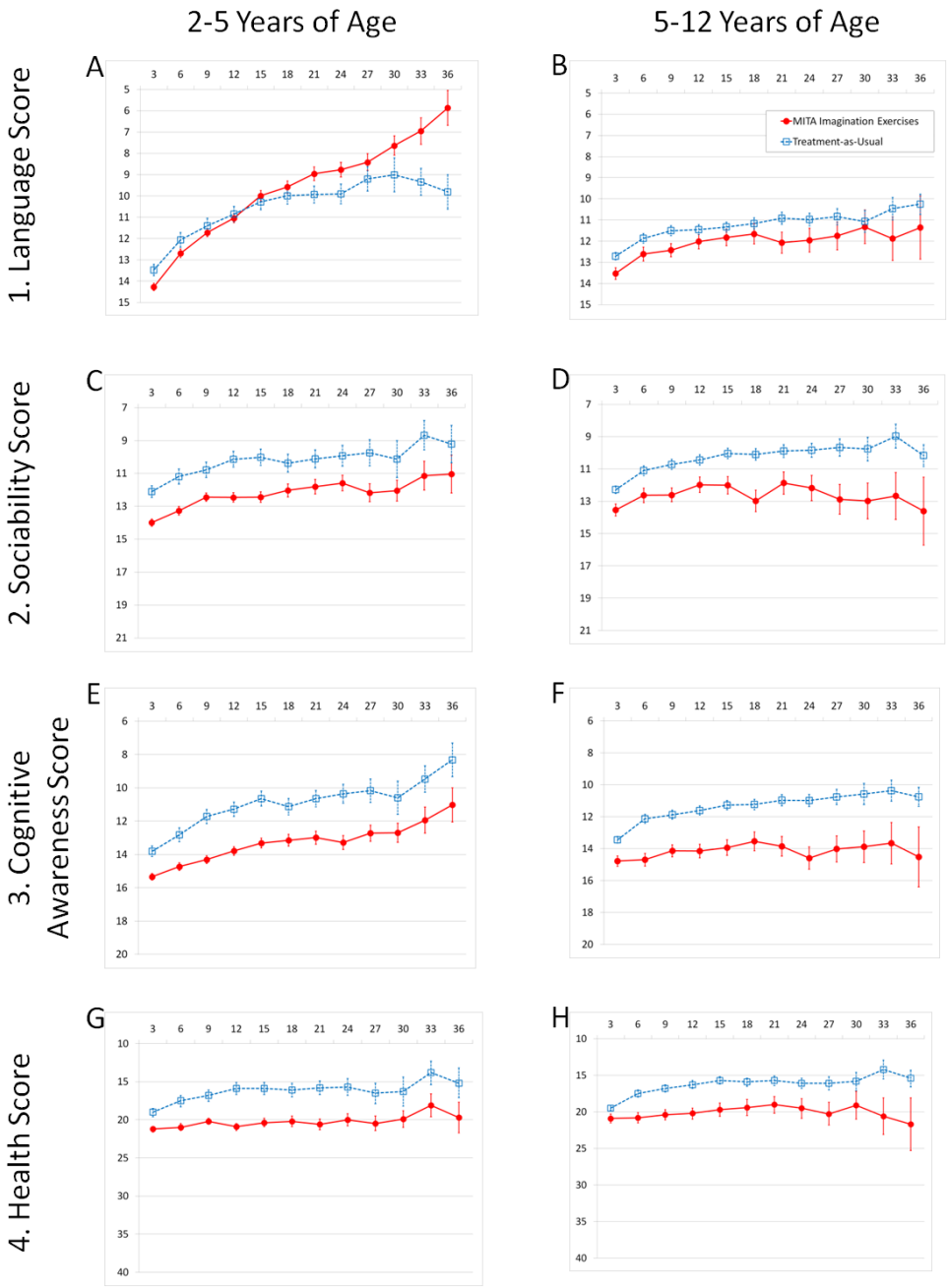

Figure 4. Longitudinal plots of subscale scores LS Means. Horizontal axis shows months from the baseline measurement (0 to 36 months). Error bars show the 95% confidence interval. To facilitate comparison between subscales, all vertical axes show 18% to 54% of their corresponding subscale' maximum available score. A lower score indicates symptoms improvement. (A) Subscale 1. Language score in younger and (B) older children. (C) Subscale 2. Sociability in younger and (D) older children. (E) Subscale 3. Cognitive awareness in younger and (F) older children. (G) Subscale 4. Health in younger and (H) older children.

To emphasize the difference in improvement between younger (2-5 YOA) and older (5-12 YOA) children groups we calculated LS mean difference between TaU and MITA at each visit, Figure 5. There was little difference in symptoms improvement between younger and older age groups in all subscales except in the Language subscale.

In the Language subscale, there was no improvement of language in MITA compared to TaU among older children (triangles and dashed line). Conversely, in the younger age group, using MITA was associated with gradual improvement of the Language score compared to control (diamonds and solid line). The difference between TaU and MITA in the younger age group attained statistical significance by visit 7 (18 months, Table S11).

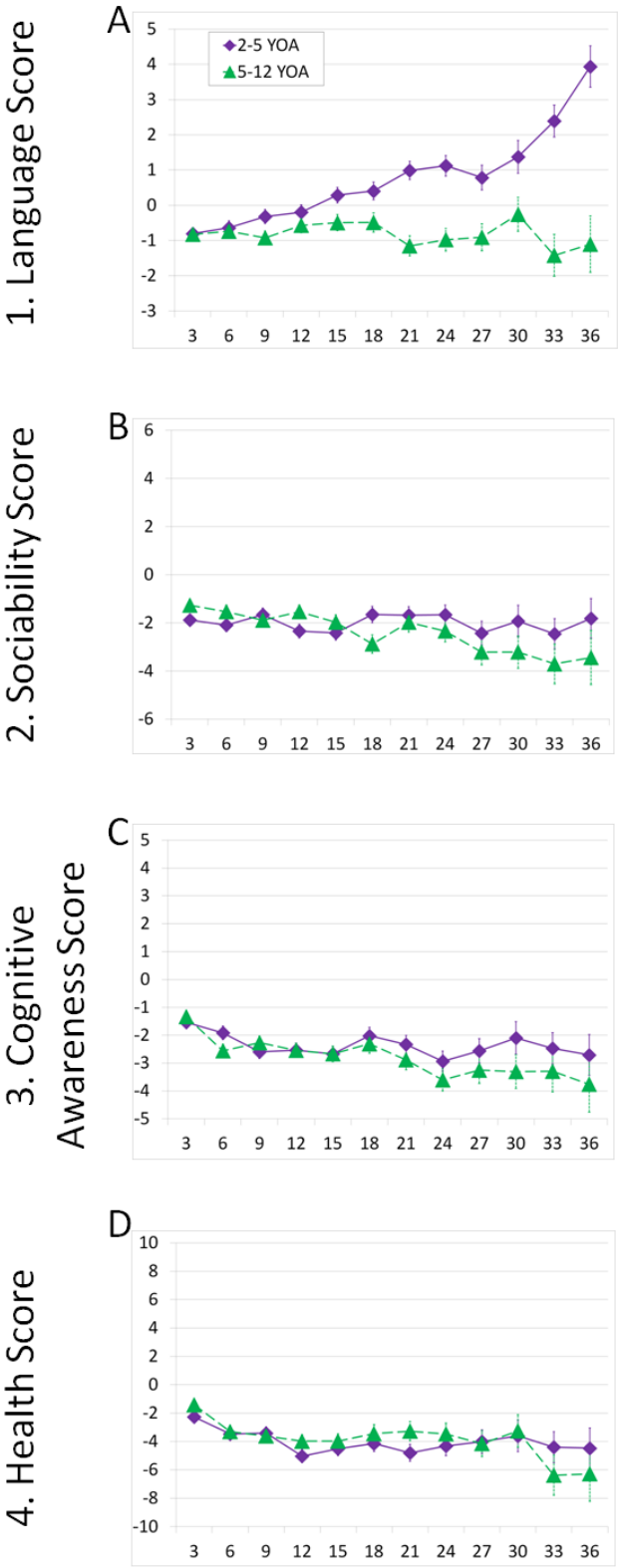

Figure 5. Longitudinal plots of contrasts between the treatment and control groups: LS Mean TaU – LS Mean MITA. Horizontal axis shows months from the baseline measurement (0 to 36 months). Error bars show SE. To facilitate comparison between subscales, all vertical axes show 30% of their corresponding subscale' maximum available score. No benefit of MITA was observed in all subscales except in the Language subscale. In the age group 2-5 YOA using MITA was associated with gradual improvement of the Language score compared to control.

	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	14.27 (0.09; 14.1 - 14.44)	13.47 (0.14; 13.2 - 13.7)	0.80 (0.16; <0.0001)	5.87 (0.42; 5.05- 6.69)	9.81 (0.41; 9.0 - 10.62)	-3.94 (0.59; <0.0001)	-8.40 (0.42; <0.0001)	-3.66 (0.41; <0.0001)
Subscale 2 Sociability	13.99 (0.12; 13.37 - 14.22)	12.11 (0.18; 11.75 - 12.47)	1.87 (0.21; <0.0001)	11.04 (0.59; 9.89 - 12.19)	9.22 (0.58; 8.08 - 10.36)	1.81 (0.82; 0.0276)	-2.95 (0.59; <0.0001)	-2.89 (0.59; 0.0086)
Subscale 3 Cognitive Awareness	15.34 (0.10; 15.44 - 15.92)	13.81 (0.16; 13.53 - 14.25)	1.53 (0.19; <0.0001)	11.02 (0.52; 10.0 - 12.05)	8.32 (0.52; 7.3 – 9.33)	2.71 (0.73; 0.0002)	-4.32 (0.53; <0.0001)	-5.49 (0.52; <0.0001)
Subscale 4 Health	21.20 (0.18; 20.92 - 21.76)	19.0 (0.28; 18.28 - 19.54)	2.26 (0.32; <0.0001)	19.7 (1.01; 17.7 - 21.7)	15.2 (1.0; 13.3 - 17.2)	4.46 (1.42; 0.0017)	-1.56 (1.03; 0.1265)	-3.76 (1.03; 0.0003)

Table 3: 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	13.53 (0.14; 13.26 - 13.81)	12.71 (0.09; 12.53 - 12.89)	0.82 (0.16; <0.0001)	11.36 (0.76; 9.86 - 12.85)	10.25 (0.25; 9.77 - 10.74)	1.10 (0.80; 0.1678)	-2.18 (0.77; 0.0046)	-2.45 (0.28; <0.0001)
Subscale 2 Sociability	13.54 (0.19; 13.17 – 13.91)	12.28 (0.12; 12.05 - 12.51)	1.26 (0.21; <0.0001)	13.61 (1.01; 11.51 - 15.72)	10.17 (0.34; 9.50 - 10.85)	3.44 (1.13; 0.0023)	0.07 (1.08; 0.9481)	-2.11 (0.35; <0.0001)
Subscale 3 Cognitive Awareness	14.78 (0.17; 14.45 - 15.11)	13.45 (0.11; 13.24 - 13.66)	1.33 (0.19; <0.0001)	14.52 (0.96; 12.65 - 16.40)	10.76 (0.31; 10.16 - 11.36)	3.76 (1.00; 0.0002)	-0.26 (0.97; 0.7888)	-2.69 (0.31; <0.0001)
Subscale 4 Health	20.9 (0.29; 20.3 - 21.5)	19.5 (0.18; 19.1 - 19.8)	1.41 (0.33; <0.0001)	21.7 (1.84; 18.1 - 25.3)	15.4 (0.59; 14.3 - 16.6)	6.27 (1.93; 0.0012)	0.82 (1.86; 0.6598)	-4.05 (0.61; <0.0001)

Table 4: 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 subscale, younger MITA children improved over the three-year period by 2.95 points (SE=0.59, p<0.0001) compared to 2.89 points (SE=0.59, p=0.0086) improvement in the TaU group, Figure 4C. 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 2.11 points (SE=0.35, p<0.0001), Figure 4D. The difference in the older MITA group relative to the TaU group at Visit 12 was 3.44 points (SE=1.13, p=0.0023).

On the Cognitive awareness subscale, younger MITA children improved over the three-year period by 4.32 points (SE=0.53, p<0.0001) compared to 5.49 points (SE=0.52, p<0.0001) in the TaU group, Figure 4E. The difference in the younger MITA group relative to the TaU group at Visit 12 was 2.71 (SE=0.73, p=0.0002). The three-year improvement in the older MITA children was not statistically significant; TaU children improved by 2.69 points (SE=0.31, p<0.0001), Figure 4F. The difference in the older MITA group relative to the TaU group at Visit 12 was 3.76 (SE=1.0, p=0.0002).

On the Health subscale, younger MITA children improvement over the three-year period was not statistically significant; TaU children improved 3.76 points (SE=1.03, p=0.0003), Figure 4G. The difference in the younger MITA group relative to the TaU group at Visit 12 was 4.46 (SE=1.42, p=0.0017). The three-year improvement in the older MITA children was not statistically significant; TaU children improved by 4.05 points (SE=0.61, p<0.0001), Figure 4H. The difference in the older MITA group relative to the TaU group at Visit 12 was 6.27 (SE=1.93, p=0.0012).

Discussion

In this report, we described data from an observational trial of tablet-based imagination exercises – *Mental Imagery Therapy for Autism* or MITA (Vyshedskiy & Dunn, 2015a) – that included 3,540 children with ASD who worked with MITA for a median duration of 520 (IQR: 384-706) days and 5,222 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 the MITA and TaU groups improved their symptoms over time in all subscales. TaU participants demonstrated either similar or greater improvement that MITA participants in all subscales except on the Language subscale in the younger group. Younger MITA participants showed 2.3-fold greater improvement in language score at the end of the trial compared to the TaU group, Figure 4A. There was no difference between MITA and TaU language improvement in the older children group, Figure 4B.

Baseline difference between the MITA and TaU groups may explain greater improvement of the TaU group

Baseline ATEC total score was 75.63 ± 22.81 in the MITA group and 60.68 ± 23.56 in the TaU group, $p<0.0001$, Table 1. MITA participants reported more severe ASD symptoms at baseline in all subscales and in both age groups. Accordingly, in order to compare participants with similar baseline measures, the propensity score analysis has been used (Schneider et al., 2007). Participants in the MITA group were matched to those in the TaU group based on all available

baseline measures: age, gender, ASD severity, and all four ATEC subscales (Language, Sociability, Cognitive awareness, and Health). The propensity score procedure significantly reduced the misbalance between TaU and MITA groups, but matching of participants was not perfect: MITA baseline ASD symptoms were more severe than TaU in all subscales in both age groups (see Visit 1, Tables 3 and 4). Multiple studies have demonstrated greater improvement on multitude of developmental measures in children with milder ASD (Fountain, Winter, & Bearman, 2012). Therefore, it is possible that greater improvement in the TaU group on the Sociability, Cognitive awareness, and Health subscales, reported here, was an artifact of imperfectly matched baseline measures.

One exception to greater improvement of the TaU group, was the Language subscale in the younger age group. Young MITA participants improved their language significantly more than TaU participants, Figure 4A. As was the case with all subscales, the MITA group had more severe language deficit at baseline even after the propensity score matching procedure. After matching, MITA group was on average more severe that TaU group by 0.80 units ($p<0.0001$) in the younger cohort and by 0.82 units ($p<0.0001$) in the older cohort (Visit 1, Tables 3 and 4). This baseline difference, biased the TaU group toward greater language improvement. In fact, a previous longitudinal epidemiological study of 2,272 children using ATEC as an outcome measure has documented greater language improvement in participants with milder ASD: children aged 2-12 years with mild ASD showed an average language improvement of 2.8 units/year; children with moderate ASD showed an average language improvement of 2.4 units/year; and children with severe ASD showed an average language improvement of 1.7

units/year (Mahapatra, Khokhlovich, et al., 2018).

Given that children with milder ASD improve their language more than children with severe ASD, the baseline difference between MITA and TaU would have been expected to favor greater language improvement in the TaU group, while we have observed exactly the opposite. Thus, the baseline difference in ASD severity between the MITA and TaU groups is unlikely to explain the observed difference in language improvement.

Language improvement mechanisms

There are three possible explanations for greater improvement of language in younger MITA-exposed children: 1) selection-bias of children or parents during trial, 2) indirect effect of MITA exercises through educating parents in the techniques of language therapy, and 3) direct effect of MITA exercises on neural networks essential for language. We discuss all three possibilities in detail below.

1. Selection-bias. Difference in language improvement between the MITA and TaU groups could be explained by selection-bias of children or parents over the course of the trial. If smarter/healthier children and more motivated parents were to continue working with MITA and remain in the trial, and weaker children/less motivated parents were to dropout from the trial, then MITA-children would have demonstrated better Language outcome compared to TaU. However, if the smarter/healthier children and more motivated parents were self-selecting to continue the trial, one would expect to see improved outcomes in all areas of children development. On the contrary, the MITA group showed better outcome only in the Language

subscale that was trained by MITA (MITA-TaU = -3.94, $p<0.0001$), Table 3 (the negative MITA-TaU indicates that the MITA group had lower score at visit 12 and therefore milder symptoms). The MITA group outcomes in all other subscales were worse than in the TaU group, Table 3 (the positive MITA-TaU indicates that the TaU group had lower score at visit 12 and therefore milder symptoms): Sociability (MITA-TaU = 1.81, $p=0.0276$), Cognitive awareness (MITA-TaU = 2.71, $p=0.0002$), and Health subscales (MITA-TaU = 4.46, $p=0.0017$). Thus, selection-bias of the MITA group for healthier and smarter children or more motivated parents was not likely to be responsible for improvements in the Language score alone.

2. Indirect effect of MITA exercises through educating parents in the techniques of language therapy. MITA caregivers could learn language therapy techniques, such as *Give me the small white woodchip* or *Give me the large black woodchip*, from MITA and then extend those techniques to everyday activities multiplying the effect of exercises many-fold. A search of the MITA listing at the app store yielded several unsolicited MITA parents' reviews, such as this one – "MITA... helps me to grab ideas from the screen and into everyday" – that supported the parents learning hypothesis. To study parents learning hypothesis further, we have solicited feedback from MITA caregivers. More than half of responders reported that they have learned some language therapy techniques from MITA exercises (unpublished observations). Moreover, literature search revealed that even short-term parents' interventions administered early have been shown to have significant effects (Landa & Kalb, 2012; Parsons, Cordier, Lee, Falkmer, & Vaz, 2019) on children language acquisition, supporting the parents-learning-from-MITA-exercises hypothesis.

3. Direct effect of MITA exercises. It is possible that imagination exercises directly trained neural networks essential for language. Association of Wernicke's and Broca's areas with language is well-known. Less common is the realization that understanding of full language depends on the lateral prefrontal cortex (LPFC). Wernicke's area primarily links words with objects (Friederici, 2011), Broca's area interprets the grammar and assigns words in a sentence to a grammatical group such as noun, verb, or preposition (Friederici, 2011), but only the LPFC can synthesize the objects from memory into a novel mental image according to grammatically imposed rules (Vyshedskiy, Dunn, & Piryatinsky, 2017; Vyshedskiy, Mahapatra, et al., 2017). This latter function may be called *imagination*, but we prefer a more specific term, *Prefrontal Synthesis* (PFS) in order to distinguish this function from other components of imagination, such as dreaming, simple memory recall, spontaneous insight, mental rotation, and integration of modifiers (Vyshedskiy, 2019). PFS is defined as voluntary juxtaposition of mental objects. On the neurological level, PFS has been hypothesized to be mediated by LPFC-driven synchronization of object-encoding neuronal ensembles (Vyshedskiy, 2019; Vyshedskiy & Dunn, 2015b). All MITA exercises were designed to train child's ability to combine mental objects.

PFS is essential for understanding sentences describing combinations of objects. E.g., the sentences "The dog bit my friend" and "My friend bit the dog" use identical words and grammar. Appreciating the misfortune of the first sentence and the humor of the second sentence depends on the LPFC ability to faithfully synthesize the two objects – the friend and the dog – into a novel mental image. Similarly, understanding of spatial prepositions such as *in, on, under, over,*

beside, in front of, behind requires a subject to synthesize several objects in front of the mind's eye. For example, the request "to put a green box {inside/behind/on top of} the blue box" requires an initial mental simulation of the scene, only after which is it possible to correctly arrange the physical objects. An inability to produce a novel mental image of the green box {inside/behind/on top of} the blue box would lead to the use of trial-and-error, which in majority of cases will result in an incorrect arrangement.

PFS completely depends on the intact LPFC and patients with damaged LPFC often lose their PFS function (Baker et al., 1996; Christoff & Gabrieli, 2000; Duncan, Burgess, & Emslie, 1995; Fuster, 2008; A. Luria, 2012; Waltz et al., 1999). Fuster calls their condition "prefrontal aphasia" (Fuster, 2008) and Luria "frontal dynamic aphasia" (A. R. Luria, 1970). Fuster explains that "although the pronunciation of words and sentences remains intact, language is impoverished and shows an apparent diminution of the capacity to 'prepositionize.' The length and complexity of sentences are reduced. There is a dearth of dependent clauses and, more generally, an underutilization of what Chomsky characterizes as the potential for recursiveness of language." (We prefer to call this condition 'PFS paralysis' since aphasia is translated from Greek as "speechless" and these patients may not experience any speech deficit.)

PFS paralysis is a known problem in individuals with ASD, commonly described as *stimulus overselectivity*, or *tunnel vision*, or *the lack of multi-cue responsivity* (Lovaas et al., 1979; Ploog, 2010; Schreibman, 1988). Many techniques used by speech language pathologists (SLP) and Applied Behavioral Analysis (ABA) therapists happen to aim at improving PFS. 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" (American Speech-Language-Hearing Association, 2016). In ABA jargon, these techniques are known as "visual-visual and auditory-visual conditional discrimination" (Axe, 2008; Eikeseth & Smith, 2013; Lowenkron, 2006; Michael, Palmer, & Sundberg, 2011), "development of multi-cue responsivity" (Lovaas et al., 1979), and "reduction of stimulus overselectivity" (Ploog, 2010). In some sense, MITA exercises can be viewed as an extreme version of language therapy with minimum vocabulary training and focusing nearly exclusively on mental integration techniques. MITA exercises can also be viewed as an extension of the "matrix training" (Frampton, Wymer, Hansen, & Shillingsburg, 2016) with minimal number of words and maximum number of word combinations. We conclude that the direct effect of MITA imagination exercises on language improvement cannot be excluded and in fact could be the most parsimonious explanation for the observed results.

Nonverbal imagination exercises

The idea to use nonverbal voluntary imagination exercises in children with developmental delay dates back to Piaget, Vygotsky, and Luria (Cole, Levitin, & Luria, 2005). In his famous twin study, Luria used educational games developed from a set of blocks to try to improve one twin's voluntary imagination. Using blocks, he designed two types of learning activities: 1) build-from-elements, and 2) build-from-model. The 'build-from-elements' activity involved building a structure per design that indicated contours of individual blocks necessary for construction. The 'build-from-model' activity indicated the overall outline of the final structure but didn't specify what blocks to use (i.e. the design did not have contours of individual blocks indicated on it).

Luria then studied ten children (5 pairs of identical twins; age 5 years). One twin of each pair followed a 2.5-month program using strategy #1 and the other using strategy #2. When tested at the end of the program, both twins in a pair were equally good at discriminating elementary figures and concentrating, but the twin following program #2 was superior in both voluntary imagination and language: he planned more, had a better sense of the relation of a block to the whole structure; program #2 twins were also *more articulate* when identifying differences between their structure and the model they were working towards (Cole et al., 2005).

The use of tablet computers significantly enhances the range and adaptability of nonverbal exercises. MITA nonverbal exercises are limitless in variations, therefore avoiding routinization. Each activity is dynamic, quickly adjusting to the child's exact ability level. All activities are disguised as games that engage children. Furthermore, many MITA verbal modules start with nonverbal levels. E.g., the *Prepositions on/under* and the *Prepositions in front/behind* games have each 30 introductory nonverbal levels, whereby objects arrangement is guided by picture-example alone; 30 intermediate levels, where objects arrangement is guided by both picture-example and verbal instruction; and 30 advanced levels, where objects arrangement is guided by verbal instruction alone. Since nonverbal exercises are easier to children with ASD (R. Dunn, Elgart, Lokshina, Faisman, Khokhlovich, et al., 2017b), MITA starts with nonverbal exercises and adds up verbal exercises slowly in order to keep the fine balance between being engaging and challenging.

Observations from ASD children support Lenneberg's critical period hypothesis

A critical period was first proposed by neurologists Wilder Penfield and Lamar Roberts in 1959

and popularized by linguist Eric H. Lenneberg in 1967 based on a few cases of childhood traumatic aphasia and hemispherectomy (Lenneberg, 1967). When the left hemisphere is surgically removed before the age of five (to treat cancer or epilepsy), patients often attain normal cognitive functions in adulthood (using the one remaining hemisphere). Conversely, removal of the left hemisphere after the age of five often results in significant impairment of recursive language and voluntary imagination (Basser, 1962; Boatman et al., 1999; Krashen & Harshman, 1972; Lenneberg, 1967; Pulsifer et al., 2004). Lenneberg's critical period hypothesis is still debated today and until now have never been tested in a large group of children.

Greater language improvement in the younger but **not** in older MITA children compared to the TaU group (Figure 5), supports the critical period hypothesis independent of the exercises' exact mode of action. 1) If imagination exercises directly trained neural networks essential for language, then they would be expected to have the same critical period as normal recursive dialogs (Romeo et al., 2018). 2) Similarly, if parents learned the techniques of language therapy by working through MITA exercises together with children, MITA also would be expected to have the same critical period as normal recursive dialogs. 3) Even if greater improvement in language was caused by selection-bias, such as greater motivation of MITA parents, our results still support the strong critical period. If we assume such selection-bias, then we should assume the same selection-bias in both the younger and older groups. However, there was no difference in language improvement between the older MITA and TaU groups. Thus, we have to conclude that even if MITA parents were more motivated than TaU parents, their greater motivation did not have any effect on language acquisition after the age of five. Whatever the mechanism of

language improvement, it worked only up to the age of five consistent with the strong critical period. To our knowledge, this is the first large epidemiological study supporting the Lenneberg's critical period hypothesis.

Strong critical period for PFS acquisition

Strong critical periods are not unusual in central nervous system development. The most famous examples include monocular deprivation (Sherman & Spear, 1982), filial imprinting in birds (Bateson, 1979), and monaural occlusion (Knudsen, Knudsen, & Esterly, 1984). Note that PFS critical period is different from other language-related critical periods, such as phoneme tuning (Kral, 2013; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992), grammar processing (Wartenburger et al., 2003), articulation control (Kim, Relkin, Lee, & Hirsch, 1997), and vocabulary acquisition (Snow & Hoefnagel-Höhle, 1978) that can be all significantly improved by training at any age (Kilgard & Merzenich, 1998; Tallal et al., 1996) and, therefore, are weak critical periods.

Less specific, more ambiguous definitions of PFS-like abilities water-down its strong critical period and undercut the analysis of language acquisition. For example, PFS ability is often rolled into one of the more general abilities such as executive function, cognition, fluid intelligence, and working memory. None of those traits have a strong critical period since they can be improved well into adulthood (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008). In addition, theory-of-mind (ToM) is often included into PFS-like abilities. Similar to PFS, ToM acquisition has a critical period: deaf children who acquire formal sign language early, are significantly better at reasoning about mental states than language-delayed deaf children (Morgan & Kegl, 2006; Pyers

& Senghas, 2009). ToM, however, improves at any age when individuals learn mental state vocabulary — particularly linguistic forms for verbs such as ‘think’ and ‘know’ (Pyers & Senghas, 2009). Therefore, ToM has a weak critical period and shall not be merged with PFS that has a strong critical period.

Mental rotation and integration of modifiers are often also grouped together with PFS since all of these three voluntary imagination processes are controlled by the LPFC. Similar to PFS, acquisition of mental rotation and integration of modifiers have critical periods (A. Martin et al., 2013; Pyers et al., 2010). However, both mental rotation and integration of modifiers can be acquired in adulthood and therefore have weak critical periods (Curtiss, 1977; Grimshaw, Adelstein, Bryden, & MacKinnon, 1998). Accordingly, for the purposes of language acquisition, PFS must be considered separately from mental rotation and integration of modifiers.

Similar to other traits with strong critical periods – monocular deprivation, filial imprinting in birds, and monaural occlusion – PFS cannot be acquired in adulthood. Its neural infrastructure has to be laid down in early childhood. Perhaps, this neural infrastructure is related to cortical functional specialization established through competition mechanisms similar to that of monocular deprivation (Ferjan Ramirez et al., 2013; Hinkley et al., 2016) and fine-tuning of long frontoposterior fibers, such as arcuate fasciculus and superior longitudinal fasciculus (Wilson et al., 2011), connecting these highly specialized cortical areas. The exact mechanism of the strong critical period for PFS acquisition remains to be determined.

Duration of the critical period

Greater language improvement in the MITA-exposed children younger than 5 years compared to TaU, but not in MITA-exposed children older than 5 years suggests that the plasticity essential for acquisition of first language reduces significantly after the age of five years. Note that the critical period in children with ASD may be shorter than in neurotypical children (Liu et al., 2016).

One of the reasons for the high prevalence of PFS paralysis and the associated lack of recursive language comprehension in ASD individuals is a deep misunderstanding of the critical period. It is not uncommon for parents to ignore their child's language delay until his/her entry into elementary school, at which time, according to the data presented herein, 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 (Webb, Jones, Kelly, & Dawson, 2014). 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. Crucially, our results imply that the underlying plasticity dramatically diminishes after the age of five and therefore even greater emphasis should be put on therapeutic intervention targeting the very first years of a child's life.

Potential issues associated with use of tablet computers for therapy administration

Giving caregivers an opportunity to administer language therapy on a tablet device comes with important warnings associated with the use of a tablet. Millions of apps are highly addictive to children (Madigan, Browne, Racine, Mori, & Tough, 2019). Most children left to their own

devices will watch YouTube for hours. A tablet device introduces an opportunity for a caregiver to leave a child in what is perceived to be a safe and productive environment. A therapeutic app carries a significant danger of being a gateway to YouTube. Having said that, we note that tablet computers are already a common staple of most families, particularly families with minimally-verbal ASD children who use tablets for communication (Lorah, Parnell, Whitby, & Hantula, 2015). Moreover, cognitively challenging apps like MITA are less addictive, as every step requires a mental effort. Once children run out of "cognitive energy" they lose their interest in MITA. Furthermore, MITA was designed to provide minimum sensory stimulation to reduce distractions and further lower its potential for addictiveness. It is also important to stress that every parent using MITA has signed a consent form that informed them of the danger of screen time for children and explained how to lock their tablet on the MITA application. Finally, in the future, MITA can be provided on a dedicated standalone device with no ability to download other apps to completely avoid the danger of exposure to YouTube and other addictive apps.

Why was the MITA effect on Language so significant?

Younger MITA-exposed children showed 2.3-fold improvement in language score at the end of the trial vs. the TaU group. This result is significantly better than reported in other trials (Dawson et al., 2010; Parsons et al., 2019). There could be several reasons for greater improvement reported herein:

1) There may exist a synergy between MITA exercises and language therapy administered by SLPs and ABA therapists. SLP and ABA techniques also aim to improve PFS, but PFS exercises are just a small part of language therapy that primarily focuses on building up the child's

vocabulary. Word comprehension is a low hanging fruit. It is easier to train and also highly appreciated by parents. Furthermore, most tests rely exclusively on a child's vocabulary to measure educational success (e.g., Peabody Picture Vocabulary Test (PPVT-4) (L. M. Dunn & Dunn, 2007), Expressive Vocabulary Test (EVT-2) (Williams, 1997)), thus encouraging therapists to focus on vocabulary training. However, such training by itself does not train PFS that is essential for understanding of spatial preposition, recursion, and complex language. Thus, the success of MITA intervention may derive from its exclusive focus on voluntary imagination and its most advanced component, PFS.

2) Even when therapists administer PFS exercises, the training is mostly verbal in nature. The intuitive verbal approach is working well in neurotypical children, but can be abstruse for nonverbal and minimally-verbal children with ASD. MITA, on the other hand, starts with nonverbal exercises that are much easier for children with ASD (R. Dunn, Elgart, Lokshina, Faisman, Khokhlovich, et al., 2017b).

3) Another hidden advantage of computerized language therapy for children with ASD could be its prosody stability. All MITA instructions are drawn from a pre-recorded library. Unlike human-given instructions, MITA verbal instructions are always pronounced with the same intonation. There are no variations in prosody. This prosody stability simplifies instruction interpretation for children who have auditory processing problems.

4) MITA observational trial was longer than an average RCT. The data reported here show significant difference in the language score between MITA and TaU only during year three, Figure 4A. If we were to stop the trial at the end of year two, we would not have detected the

effect of MITA. In retrospect, the long time necessary for detecting effects of imagination exercises should have been expected. PFS is an internal mental ability and it does not immediately manifest itself to caregivers and psychometricians (Vyshedskiy, 2019). There are currently no tests that focus on PFS acquisition. Caregivers and psychometricians commonly assess expressive language that depends on PFS, but is also influenced by many other factors, such as the control of speech apparatus. It may take several years before an initially nonverbal or minimally verbal child expresses his/her PFS skills through expressive language.

5) The most significant challenge of conventional therapy is a substantial cost that significantly reduces therapists availability to most families (Peters-Scheffer, Didden, Korzilius, & Matson, 2012). The free MITA application, on the other hand, is always available and its everyday use does not increase the financial burden.

6) Unlike a human therapist, MITA is readily available for download and use within minutes. As a result, parents can initiate MITA exercises at a very early age, even before the official ASD diagnosis. Most MITA parents start MITA around the age of two and indicate their diagnosis as "suspected ASD" at registration, they change their diagnosis to ASD months later.

7) To receive optimal therapy, children have to develop a certain degree of connection with a therapist (McIntyre & Barton, 2010). In a field in which frequent rotation of therapists is a norm, a lot of time is wasted on the initial therapist-child bridge-building (Zimmermann et al., 2019). MITA, on the other hand, has only a minimal break-in period.

8) Children often miss a scheduled therapy session due to travel, sickness, tantrums, or lack of

focus (Carr et al., 2016). Once a session has been missed, a therapist may not be available for a make-up lesson. One of the benefits of MITA is that parents can administer MITA anytime when they feel that children are in a good mood and receptive to therapy.

9) When parents delegate all therapy to professionals, they may not participate in therapy sessions, and, as a result, may never learn language therapy techniques. Consequently, the most valuable time that could have been used for parent-child communication is missed. MITA inadvertently forces parents to work with their children through language exercises, therefore promoting parents' learning of language therapy techniques.

10) Finally, MITA gives parents hope and it is this hope that helps parents to motivate their children and persist with language therapy. By seeing their children solving puzzles, parents become more confident of their children. As one parent wrote in an unsolicited review: "My son displays an intellectual capability, which I thought for a long time was missing."

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 a randomized controlled trial (RCT). Prior to conducting the MITA study, we have submitted the proposals for a therapist-administered RCT of voluntary imagination intervention to many potential funding agencies. 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 biweekly telephone calls to encourage app use (Whitehouse et al., 2017): 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 (Whitehouse et al., 2017). As the minimal length of a voluntary imagination intervention RCT is likely to exceed two years (Dawson et al., 2010; Smith, Groen, & Wynn, 2000), participant dropout becomes the major issue. This high attrition rate introduces multiple selection biases that devalue RCT ability to demonstrate causality and essentially makes it no better than an observational trial. The reported self-funded observational trial is the best study we could conduct without external funding.

Another disadvantage of the low-cost geographically diverse observational trials is their reliance on parent-reported outcome measures. There is an understanding within 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 (Scattone, Raggio, & May, 2011). 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 (Al Backer, 2016; Charman et al., 2004; Geier et al., 2013; Jarusiewicz,

2002; Klaveness et al., 2013; Magiati et al., 2011; Mahapatra, Khokhlovich, et al., 2018; Whitehouse et al., 2017), assessing participants on four diverse scales of development (Language, Sociability, Cognitive awareness, and Health) and administered at regular 3-month intervals.

Our results should be treated cautiously, as less motivated families may not be able to commit themselves to long-term therapy administration and families without technical backgrounds may find their experience with MITA less intuitive. Further validation of MITA exercises is necessary to understand its efficacy within the diverse autism population.

Conclusions

Five major conclusions follow up from this study. First, in children with ASD, the plasticity essential for acquisition of the first language is significantly reduced after the age of five years consistent with Lenneberg's critical period hypothesis (Lenneberg, 1967). Second, voluntary imagination is an essential component of full language and exercises training imagination are an indispensable component of language therapy. Third, the minimal necessary duration of a clinical trial investigating the effect of imagination exercises is three years. Fourth, some caregivers are capable of administering tablet-based exercises to their children consistently over many years. Fifth, parent-administered and 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 to translate MITA to

multiple languages. The major strength of this study is the large number of long-term participants. The most obvious limitation is that this study observational design cannot definitively prove causality since not all confounders can be adjusted appropriately. We conclude that MITA exercises warrant further investigation in a randomized controlled study (RCT).

Acknowledgments

We wish to thank Dr. Noah Greifer for assistance with statistical analysis, Danielle Abate for assistance with manuscript preparation and Yulia Dumov for the design of children's 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

Al Backer, N. B. (2016). Correlation between Autism Treatment Evaluation Checklist (ATEC) and Childhood Autism Rating Scale (CARS) in the evaluation of autism spectrum disorder. *Sudanese Journal of Paediatrics*, *16*(1), 17.

American Speech-Language-Hearing Association. (2016). *Scope of practice in speech-language pathology*.

Axe, J. B. (2008). Conditional discrimination in the intraverbal relation: A review and recommendations for future research. *The Analysis of Verbal Behavior*, *24*(1), 159–174.

Baker, S. C., Rogers, R. D., Owen, A. M., Frith, C. D., Dolan, R. J., Frackowiak, R. S. J., & Robbins, T. W. (1996). Neural systems engaged by planning: A PET study of the Tower of London task. *Neuropsychologia*, *34*(6), 515–526.

Basser, L. S. (1962). Hemiplegia of early onset and the faculty of speech with special reference to the effects of hemispherectomy. *Brain*, *85*(3), 427–460.

Bateson, P. (1979). Brief exposure to a novel stimulus during imprinting in chicks and its influence on subsequent preferences. *Animal Learning & Behavior*, *7*(2), 259–262.

Bick, J., Zhu, T., Stamoulis, C., Fox, N. A., Zeanah, C., & Nelson, C. A. (2015). Effect of early institutionalization and foster care on long-term white matter development: A randomized clinical trial. *JAMA Pediatrics*, *169*(3), 211–219.

Boatman, D., Freeman, J., Vining, E., Pulsifer, M., Miglioretti, D., Minahan, R., … McKhann, G. (1999). Language recovery after left hemispherectomy in children with late-onset seizures. *Annals of Neurology*, *46*(4), 579–586.

Carr, T., Shih, W., Lawton, K., Lord, C., King, B., & Kasari, C. (2016). The relationship between treatment attendance, adherence, and outcome in a caregiver-mediated intervention for low-resourced families of young children with autism spectrum disorder. *Autism*, *20*(6), 643–652.

Charman, T., Howlin, P., Berry, B., & Prince, E. (2004). Measuring developmental progress of children with autism spectrum disorder on school entry using parent report. *Autism*, *8*(1), 89–100.

Cheng, Q., Roth, A., Halgren, E., & Mayberry, R. I. (2019). Effects of early language deprivation on brain connectivity: Language pathways in deaf native and late first-language learners of American Sign Language. *Frontiers in Human Neuroscience*, *13*, 320.

Christoff, K., & Gabrieli, J. D. (2000). The frontopolar cortex and human cognition: Evidence for a rostrocaudal hierarchical organization within the human prefrontal cortex. *Psychobiology*, *28*(2), 168–186.

Cole, M., Levitin, K., & Luria, A. R. (2005). *Autobiography of Alexander Luria: A Dialogue with the Making of Mind* (1 edition). New York: Psychology Press.

Curtiss, S. (1977). *Genie: A Psycholinguistic Study of a Modern-Day"" Wild Child""(Perspectives in*. Retrieved from http://globalloveins.com/book18/88012.pdf

Dawson, G., Rogers, S., Munson, J., Smith, M., Winter, J., Greenson, J., … Varley, J. (2010). Randomized, controlled trial of an intervention for toddlers with autism: The Early Start Denver Model. *Pediatrics*, *125*(1), e17–e23.

Dube, W. V., Farber, R. S., Mueller, M. R., Grant, E., Lorin, L., & Deutsch, C. K. (2016). Stimulus overselectivity in autism, Down syndrome, and typical development. *American Journal on Intellectual and Developmental Disabilities*, *121*(3), 219–235.

Duncan, J., Burgess, P., & Emslie, H. (1995). Fluid intelligence after frontal lobe lesions. *Neuropsychologia*, *33*(3), 261–268.

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody picture vocabulary test*. Pearson Assessments.

Dunn, R., Elgart, J., Lokshina, L., Faisman, A., Khokhlovich, E., Gankin, Y., & Vyshedskiy, A. (2017a). 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*(217). https://doi.org/10.4172/2165-7890.1000217

Dunn, R., Elgart, J., Lokshina, L., Faisman, A., Khokhlovich, E., Gankin, Y., & Vyshedskiy, A. (2017b). Comparison of performance on verbal and nonverbal multiple-cue responding

tasks in children with ASD. *Autism Open Access*, *7*, 218. https://doi.org/10.4172/2165-7890.1000218

Dunn, R., Elgart, J., Lokshina, L., Faisman, A., Waslick, M., Gankin, Y., & Vyshedskiy, A. (2017). Tablet-Based Cognitive Exercises as an Early Parent-Administered Intervention Tool for Toddlers with Autism—Evidence from a Field Study. *Clinical Psychiatry*, *3*(1). https://doi.org/10.21767/2471-9854.100037

Eikeseth, S., & Smith, D. P. (2013). An analysis of verbal stimulus control in intraverbal behavior: Implications for practice and applied research. *The Analysis of Verbal Behavior*, *29*(1), 125–135.

Ferjan Ramirez, N., Leonard, M. K., Torres, C., Hatrak, M., Halgren, E., & Mayberry, R. I. (2013). Neural language processing in adolescent first-language learners. *Cerebral Cortex*, *24*(10), 2772–2783.

Fombonne, E. (2003). Epidemiological surveys of autism and other pervasive developmental disorders: An update. *Journal of Autism and Developmental Disorders*, *33*(4), 365–382.

Fountain, C., Winter, A. S., & Bearman, P. S. (2012). Six developmental trajectories characterize children with autism. *Pediatrics*, *129*(5), e1112–e1120.

Frampton, S. E., Wymer, S. C., Hansen, B., & Shillingsburg, M. A. (2016). The use of matrix training to promote generative language with children with autism. *Journal of Applied Behavior Analysis*, *49*(4), 869–883.

Friederici, A. D. (2011). The brain basis of language processing: From structure to function. *Physiological Reviews*, *91*(4), 1357–1392.

Fuster, J. (2008). *The Prefrontal Cortex, Fourth Edition* (4 edition). Amsterdam; Boston: Academic Press.

Geier, D. A., Kern, J. K., & Geier, M. R. (2013). A comparison of the Autism Treatment Evaluation Checklist (ATEC) and the Childhood Autism Rating Scale (CARS) for the quantitative evaluation of autism. *Journal of Mental Health Research in Intellectual Disabilities*, *6*(4), 255–267.

Grimshaw, G. M., Adelstein, A., Bryden, M. P., & MacKinnon, G. E. (1998). First-language acquisition in adolescence: Evidence for a critical period for verbal language development. *Brain and Language*, *63*(2), 237–255.

Hinkley, L. B., Marco, E. J., Brown, E. G., Bukshpun, P., Gold, J., Hill, S., … Barkovich, A. J. (2016). The contribution of the corpus callosum to language lateralization. *Journal of Neuroscience*, *36*(16), 4522–4533.

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences*, *105*(19), 6829–6833.

Jarusiewicz, B. (2002). Efficacy of Neurofeedback for Children in the Autistic Spectrum: A Pilot Study. *Journal of Neurotherapy*, *6*(4), 39–49. https://doi.org/10.1300/J184v06n04_05

Kilgard, M. P., & Merzenich, M. M. (1998). Plasticity of temporal information processing in the primary auditory cortex. *Nature Neuroscience*, *1*(8), 727.

Kim, K. H., Relkin, N. R., Lee, K.-M., & Hirsch, J. (1997). Distinct cortical areas associated with native and second languages. *Nature*, *388*(6638), 171–174.

Klaveness, J., Bigam, J., & Reichelt, K. L. (2013). The varied rate of response to dietary intervention in autistic children. *Open Journal of Psychiatry*, *3*(02), 56.

Knudsen, E. I., Knudsen, P. F., & Esterly, S. D. (1984). A critical period for the recovery of sound localization accuracy following monaural occlusion in the barn owl. *Journal of Neuroscience*, *4*(4), 1012–1020.

Kral, A. (2013). Auditory critical periods: A review from system's perspective. *Neuroscience*, *247*, 117–133.

Krashen, S., & Harshman, R. (1972). Lateralization and the critical period. *The Journal of the Acoustical Society of America*, *52*(1A), 174–174.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*(5044), 606–608.

Landa, R. J., & Kalb, L. G. (2012). Long-term outcomes of toddlers with autism spectrum disorders exposed to short-term intervention. *Pediatrics*, *130*(Supplement 2), S186–S190.

Lenneberg, E. H. (1967). The biological foundations of language. *Hospital Practice*, *2*(12), 59–67.

Liu, X., Han, D., Somel, M., Jiang, X., Hu, H., Guijarro, P., … Ely, J. J. (2016). Disruption of an evolutionarily novel synaptic expression pattern in autism. *PLoS Biology*, *14*(9).

Lorah, E. R., Parnell, A., Whitby, P. S., & Hantula, D. (2015). A systematic review of tablet computers and portable media players as speech generating devices for individuals with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *45*(12), 3792–3804.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., … Rutter, M. (2000). 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*, *30*(3), 205–223.

Lovaas, O. I., Koegel, R. L., & Schreibman, L. (1979). Stimulus overselectivity in autism: A review of research. *Psychological Bulletin*, *86*(6), 1236–1254.

Lowenkron, B. (2006). Joint control and the selection of stimuli from their description. *The Analysis of Verbal Behavior*, *22*(1), 129–151.

Luria, A. (2012). *Higher cortical functions in man*. Springer Science & Business Media.

Luria, A. R. (1970). Traumatic aphasia. Mouton. *The Hague*.

Madigan, S., Browne, D., Racine, N., Mori, C., & Tough, S. (2019). Association between screen time and children's performance on a developmental screening test. *JAMA Pediatrics*, *173*(3), 244–250.

Magiati, I., Moss, J., Yates, R., Charman, T., & Howlin, P. (2011). Is the Autism Treatment Evaluation Checklist a useful tool for monitoring progress in children with autism spectrum disorders? *Journal of Intellectual Disability Research*, *55*(3), 302–312. https://doi.org/10.1111/j.1365-2788.2010.01359.x

Mahapatra, S., Khokhlovich, E., Martinez, S., Kannel, B., Edelson, S. M., & Vyshedskiy, A. (2018). Longitudinal Epidemiological Study of Autism Subgroups Using Autism Treatment Evaluation Checklist (ATEC) Score. *Autism and Developmental Disorders*, *1*(12). https://doi.org/10.1007/s10803-018-3699-2

Mahapatra, S., Vyshedsky, D., Martinez, S., Kannel, B., Braverman, J., Edelson, S. M., & Vyshedskiy, A. (2018). Autism Treatment Evaluation Checklist (ATEC) norms: A "growth chart" for ATEC score changes as a function of age. *Children*, *5*(2). https://doi.org/10.3390

Martin, A. J. (2009). *Does age of language acquisition affect the relation between American sign language and mental rotation?* (UNIVERSITY OF MINNESOTA). Retrieved from http://conservancy.umn.edu/handle/11299/57256

Martin, A., Senghas, A., & Pyers, J. (2013). Age of acquisition effects on mental rotation: Evidence from Nicaraguan sign language. *BUCLD 37: Proceedings of the 37th Boston*

University Conference on Language Development, 241–250. Retrieved from http://www.academia.edu/download/37807045/MartinSenghasPyers2013.pdf

McIntyre, L. L., & Barton, E. E. (2010). Early childhood autism services: How wide is the research to practice divide? *Behavioral Development Bulletin*, *16*(1), 34.

Michael, J., Palmer, D. C., & Sundberg, M. L. (2011). The multiple control of verbal behavior. *The Analysis of Verbal Behavior*, *27*(1), 3–22.

Morgan, G., & Kegl, J. (2006). Nicaraguan sign language and theory of mind: The issue of critical periods and abilities. *Journal of Child Psychology and Psychiatry*, *47*(8), 811–819.

Parsons, D., Cordier, R., Lee, H., Falkmer, T., & Vaz, S. (2019). A Randomised Controlled Trial of an Information Communication Technology Delivered Intervention for Children with Autism Spectrum Disorder Living in Regional Australia. *Journal of Autism and Developmental Disorders*, *49*(2), 569–581.

Pearson, J. (2019). The human imagination: The cognitive neuroscience of visual mental imagery. *Nature Reviews Neuroscience*, 1–11.

Peters-Scheffer, N., Didden, R., Korzilius, H., & Matson, J. (2012). Cost comparison of early intensive behavioral intervention and treatment as usual for children with autism spectrum disorder in the Netherlands. *Research in Developmental Disabilities*, *33*(6), 1763–1772.

Ploog, B. O. (2010). Stimulus overselectivity four decades later: A review of the literature and its implications for current research in autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *40*(11), 1332–1349.

Pulsifer, M. B., Brandt, J., Salorio, C. F., Vining, E. P., Carson, B. S., & Freeman, J. M. (2004). The cognitive outcome of hemispherectomy in 71 children. *Epilepsia*, *45*(3), 243–254.

Pyers, J. E., & Senghas, A. (2009). Language promotes false-belief understanding: Evidence from learners of a new sign language. *Psychological Science*, *20*(7), 805–812.

Pyers, J. E., Shusterman, A., Senghas, A., Spelke, E. S., & Emmorey, K. (2010). Evidence from an emerging sign language reveals that language supports spatial cognition. *Proceedings of the National Academy of Sciences*, *107*(27), 12116–12120.

Rimland, B., & Edelson, S. (1999). Autism Research Institute. *Autism Treatment Evaluation Checklist (ATEC)*.

Romeo, R. R., Segaran, J., Leonard, J. A., Robinson, S. T., West, M. R., Mackey, A. P., … Gabrieli, J. D. (2018). Language exposure relates to structural neural connectivity in childhood. *Journal of Neuroscience*, *38*(36), 7870–7877.

Scattone, D., Raggio, D. J., & May, W. (2011). Comparison of the vineland adaptive behavior scales, and the bayley scales of infant and toddler development. *Psychological Reports*, *109*(2), 626–634.

Schneider, B., Carnoy, M., Kilpatrick, J., Schmidt, W. H., & Shavelson, R. J. (2007). *Estimating*

causal effects using experimental and observational design. American Educational & Reseach Association.

Schreibman, L. (1988). Diagnostic features of autism. *Journal of Child Neurology*, *3*(1 suppl), S57–S64.

Sherman, S. M., & Spear, P. D. (1982). Organization of visual pathways in normal and visually deprived cats. *Physiological Reviews*, *62*(2), 738–855.

Skeide, M. A., Brauer, J., & Friederici, A. D. (2015). Brain functional and structural predictors of language performance. *Cerebral Cortex*, *26*(5), 2127–2139.

Smith, T., Groen, A. D., & Wynn, J. W. (2000). Randomized trial of intensive early intervention for children with pervasive developmental disorder. *American Journal on Mental Retardation*, *105*(4), 269–285.

Snow, C. E., & Hoefnagel-Höhle, M. (1978). The critical period for language acquisition: Evidence from second language learning. *Child Development*, 1114–1128.

Tallal, P., Miller, S. L., Bedi, G., Byma, G., Wang, X., Nagarajan, S. S., … Merzenich, M. M. (1996). Language comprehension in language-learning impaired children improved with acoustically modified speech. *Science*, *271*(5245), 81–84.

Vyshedskiy. (2019). Neuroscience of imagination and implications for hominin evolution. *Journal of Current Neurobiology*. https://doi.org/10.31234/osf.io/skxwc

Vyshedskiy, A., DuBois, M., Mugford, E., Piryatinsky, I., Radi, K., Braverman, J., & Maslova, V. (2020). Novel Linguistic Evaluation of Prefrontal Synthesis (LEPS) test measures prefrontal synthesis acquisition in neurotypical children and predicts high-functioning versus low-functioning class assignment in individuals with autism. *Applied Neuropsychology: Child*. https://doi.org/10.1080/21622965.2020.1758700

Vyshedskiy, A., & Dunn, R. (2015a). Mental Imagery Therapy for Autism (MITA)-An Early Intervention Computerized Brain Training Program for Children with ASD. *Autism Open Access*, *5*(1000153), 2.

Vyshedskiy, A., Dunn, R., & Piryatinsky, I. (2017). Neurobiological mechanisms for nonverbal IQ tests: Implications for instruction of nonverbal children with autism. *Research Ideas and Outcomes*, *3*, e13239. https://doi.org/10.3897/rio.3.e13239

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

Vyshedskiy, & Dunn, R. (2015b). Mental synthesis involves the synchronization of independent neuronal ensembles. *Research Ideas and Outcomes*, *1*, e7642.

Waltz, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Mishkin, F. S., de Menezes Santos, M., … Miller, B. L. (1999). A system for relational reasoning in human prefrontal cortex. *Psychological Science*, *10*(2), 119–125.

Wartenburger, I., Heekeren, H. R., Abutalebi, J., Cappa, S. F., Villringer, A., & Perani, D. (2003). Early setting of grammatical processing in the bilingual brain. *Neuron*, *37*(1), 159–170.

Webb, S. J., Jones, E. J. H., Kelly, J., & Dawson, G. (2014). The motivation for very early intervention for infants at high risk for autism spectrum disorders. *International Journal of Speech-Language Pathology*, *16*(1), 36–42. https://doi.org/10.3109/17549507.2013.861018

Whitehouse, A. J., Granich, J., Alvares, G., Busacca, M., Cooper, M. N., Dass, A., … others. (2017). A randomised controlled trial of an iPad-based application to complement early behavioural intervention in Autism Spectrum Disorder. *Journal of Child Psychology and Psychiatry*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/jcpp.12752/full

Williams, K. T. (1997). Expressive vocabulary test second edition (EVT™ 2). *J. Am. Acad. Child Adolesc. Psychiatry*, *42*, 864–872.

Wilson, S. M., Galantucci, S., Tartaglia, M. C., Rising, K., Patterson, D. K., Henry, M. L., … Gorno-Tempini, M. L. (2011). Syntactic processing depends on dorsal language tracts. *Neuron*, *72*(2), 397–403.

Zimmermann, D., Lutz, W., Reiser, M., Boyle, K., Schwartz, B., Schilling, V. N., … Rubel, J. A. (2019). What happens when the therapist leaves? The impact of therapy transfer on the therapeutic alliance and symptoms. *Clinical Psychology & Psychotherapy*, *26*(1), 135–145.

Supplementary Information is linked to the online version of the paper.