The effect of enactment on working memory in children on the autism spectrum (ASD) and typically developing children (TD)

Student: A. Reimes, s4538293

First assessor: Dr. S. M. Field
Second assessor: Dr. W. J. Post

University of Groningen
Faculty of Behavioural and Social Sciences
Mastertrack Orthopedagogy
December 2024

Wordcount: 7823

Abstract

Title: The effect of enactment on working memory in children on the autism spectrum (ASD) and typically developing children (TD). **Problem:** ASD children often face challenges in their working memory, compared to TD children, which may hinder their academic performance and following the teacher's instruction. However, as far as known, there is limited research on how using enactment can improve working memory performance and thereby could be effectively implemented to support ASD children in the mainstream classroom. **Research questions:** This study aims to give insight into the facilitating effect of enactment on working memory performance for ASD and TD children and explores the role of age on this relationship. **Method:** In this quantitative study, a mixed (between and within subject)design has been used. The Span task was used to measure the working memory performance of both ASD and TD children. The effect of enactment and the role of age was measured by using a 2x2 Repeated Measures ANOVA and a Repeated Measures ANCOVA. **Results:** A significant effect was found in both group and instruction type on the working memory capacity of ASD and TD children, with a benefit for TD children, however the statistical test may have produced unreliable findings due to normality violations. Additionally, age appears to have an extremely minimal impact on the working memory performance, whereas group and interaction between instruction type and group significantly influence this. **Conclusions:** Several explanations are given for this limited effect. Further research on the influence of enactment is recommended.

Samenvatting

Titel: Het effect van enactment op het werkgeheugen van kinderen met een autismespectrumstoornis (ASS) en typisch ontwikkelende kinderen (TD). **Probleemstelling:** ASS-kinderen hebben in vergelijking met TD-kinderen vaak problemen met hun werkgeheugen, wat hun academische prestaties en het volgen van instructies van de leerkracht kan belemmeren. Voor zover bekend is er echter beperkt onderzoek gedaan naar hoe het gebruik van enactment de prestaties van het werkgeheugen van kinderen kan verbeteren en daardoor effectief kan worden toegepast om ASS-kinderen te ondersteunen in het reguliere onderwijs. **Vraagstelling:** Deze studie heeft als doel om inzicht te krijgen in het faciliterende effect van enactment op werkgeheugenprestaties voor ASS- en TD kinderen, waarbij de rol van leeftijd op deze relatie is onderzocht. **Methode:** In deze kwantitatieve studie is gebruikgemaakt van een mixed (between and within subject)design. De Span-taak werd gebruikt om werkgeheugenprestaties van zowel ASS- als TD kinderen te meten. Het effect van enactment en de rol van leeftijd werd gemeten met behulp van een 2x2 Repeated Measures ANOVA en een Repeated Measures ANCOVA. **Resultaten:** Er werd een significant effect gevonden in zowel de groep als het type instructie op de werkgeheugencapaciteit van ASS en TD kinderen, met een voordeel voor TD kinderen. De statistische test kan echter onbetrouwbare resultaten hebben opgeleverd vanwege normaliteitsschendingen. Daarnaast blijkt leeftijd een extreem lage impact te hebben op de werkgeheugenprestatie, terwijl groep en interactie tussen instructietype en groep dit significant beïnvloeden. **Conclusies:** Er worden verschillende verklaringen gegeven voor dit beperkte effect. Verder onderzoek naar de invloed van enactment wordt aanbevolen.

Table of contents

Introduction

"Autism spectrum disorder (ASD) refers any one of a group of disorders with an onset typically occurring during the preschool years and characterized by difficulties with social communication and social interaction along with restricted and repetitive patterns in behaviors, interests, and activities" (American Psychiatric Association, 2022). The fifth edition of the Diagnostic and Statistical Manual of Mental Disorder (DSM-V) (American Psychiatric Association, 2013) describes an autism spectrum disorder as a neurodevelopmental disorder. Its characteristic behavior can be seen early in the development of a child (Matson et al., 2009). The prevalence of the ASD diagnosis has dramatically increased in the last years (Eric, 2020). The core features of ASD are social communication deficits and repetitive and unusual sensory–motor behaviors (Lord et al., 2018). The functioning levels vary among children and ASD children may encounter challenges in different developmental stages and contexts (Lobar, 2016). Abnormalities in sensory information processing can be a barrier in education (Kern et al., 2006). In the classroom, ASD children can develop behavioral problems and experience difficulties in processing verbal instructions from the teacher. Therefore, these children are more likely to underachieve academically compared to typically developing children (TD) (Ashburner, Ziviani & Rodger, 2008).

The Salamanca Statement states children with special needs and disabilities have the rights and should be able to join general schools where educational needs are adapted (UNESCO, 1994). A positive aspect of inclusive education is that it develops opportunities to experience meaningful social interactions and relationships with peers (Freitag & Dunsmuir, 2015). A key factor to successful inclusion is a teacher's disposition and behavior (Burack, Root & Zigler, 1997). Therefore, teachers can experience challenges in helping ASD children manage traits related to this disorder, especially for children who are low-functioning, in the mainstream classroom (Kucharczyk et al., 2015). In addition to developing positive attitudes among teachers about inclusive education (Bolourian et al., 2021), there is a need from practitioners for strategies for inclusive education and changing the learning environment of ASD children (Petersson-Bloom & Holmqvist, 2022). However, supporting ASD children can be difficult due to the fact ASD severity varies widely. Also because ASD children demonstrate several challenging behaviors and lower abilities, such as working memory, which functions on average less well for ASD children compared with TD children (Tse et al., 2019). It is important to overcome these deficits with an intervention to improve the likelihood of their academic success (Alloway, 2006). Through appropriate education and meeting the needs of ASD children, differences within individuals in the classroom can be reduced, which benefits inclusive education (Lynch & Irvine, 2009).

Working memory is important in children's development and their academic performance (Gathercole, 2004). It is a system in the brain that holds a limited amount of information ready for processing (Oberauer, 2002) and is crucial in developing general cognitive abilities (Munakata, Morton & O'Reilly, in press). Reasoning and learning are examples of complete tasks in which working memory is seen as necessary to keep things in mind (Baddeley, 2010). Working memory tasks can be used to measure the mental resources being shared between both processing and storage components (Towse, Hitch & Hutton, 1998).

The central executive system plays a key role in driving working memory capacity. It is driven by the phonological loop, the visual-spatial sketchpad and the episodic buffer (Baddeley, 2003a). This system is responsible for directing, dividing and switching attention (Kane & Engle, 2003) and regulates and allocates attention resources (Baddeley, 2012). In ASD children there has been found an impaired function of the central executive system (Xie et al., 2024). The episodic buffer, one component of the working memory which binds information and knowledge into a coherent whole (Baddeley, 2012), shows a reduced ability to integrate different types of information in ASD children (Lecouvey et al., 2015).

Working memory performance starts early in children's development and improves when they grow older (Brockmole & Logie, 2013). From around the age of 2 years old working memory begins to function (Diamond, 2013). The short-term memory capabilities of children improve between the ages of 4 and 15, whereas the visual-short term memory grows steadily between 5 and 11 years old (Gathercole, 1998). Due to maturation of brain regions during development, children are better able to retain and process information when they grow older (Diamond, 2013). When reaching the age of 11 years, the working memory capacity of children is similar to the working memory capacity of adults (Gathercole, 1998).

There are different ways to implement strategies in education which can improve opportunities for ASD children to follow the teachers' instructions correctly and participate in mainstream classrooms. Literature shows enactment has a positive effect on the working memory for instructions in ASD children (Wang et al., 2022). Enactment is a type of action-based processing which can increase memory performance (Saltz, 1988). Enactment means the given instructions are conducted by physical actions (Wang et al., 2024). "The enactment effect" refers to the phenomenon whereby people better remember instructions or information if they physically act them out, as opposed to only reading or hearing about them. The brain registers the information more in detail (Klein & Saltz, 1976). Actively encoded material is less susceptible to forgetting (Spranger, Schatz & Knopf, 2008). Learning by self-performing actions is more efficient according to the "enactment effect" in long-term memory for both ASD children (Grainger et al., 2014) and TD children

(Engelkamp, 1998). Teachers can apply enactment in the classroom by giving student lessons in which assignments need to be physically carried out. Implementing enactment tasks in education therefore might improve the overall cognitive function of ASD children, including their working memory. It promotes children's self-regulation, attention and motivation (Summers & Craik, 1994). Following multi-step instructions and switching between tasks at school is also more easily (Dawson et al., 2010).

Research purpose- and questions

This study aims to gain insight into the facilitating effect of the action-based processing type enactment on working memory capacity (WM) for instructions in children on the autism spectrum (ASD) and typically developing children (TD), and how it varies between both groups. Additionally, the influence of age on the relationship between enactment and working memory capacity for ASD- and TD children will be studied. In total 2 research questions will be answered in this study. The first research question is: *Does the action-based processing type enactment improve working memory capacity for instructions differently for children on the autism spectrum (ASD) and typically developing children (TD), compared to traditional verbal instructions?*

The function of the central executive system affects the enactment effect in working memory for instructions in TD children (Waterman et al., 2017). ASD children show an impaired function of the central executive system and usually perform less on tasks involving working memory (Xie et al., 2024). Based on this, there are the hypothesis:

H_0: There is an interaction between group (ASD vs. TD) and condition (enactment vs. hearing) on working memory performance, with ASD children performing similar on the Span task compared to TD children in the same enactment condition.

H_A: There is an interaction between group (ASD vs. TD) and condition (enactment vs. hearing) on working memory performance, with TD children perform best on the Span task and ASD children show comparatively lower performance on the Span task in the same enactment condition.

The second research question is*: What is the influence of age on the relationship between the action-based processing type enactment and working memory capacity for instructions in children on the autism spectrum (ASD) and typically developing children (TD)?*

Research indicates that working memory improves when children grow older (Brockmole & Logie, 2013). It is seen that from the age of 11 a working memory level equivalent to that of an adult is reached (Gathercole, 1998). Based on this, these are the hypothesis:

H_0: After controlling for age, there is no significant difference in working memory performance between ASD- and TD children in the enactment condition, with older children in the ASD group performing similar on the Span task compared to TD children in either condition.
H_A: After controlling for age, there is a significant difference in working memory performance between ASD- and TD children in the enactment condition, with older children in the TD group performing better on the Span task than ASD children in either condition.

Research methodology

Design

This is a quantitative study, making use of existing data from the study of Xie et al. (2024). The originating study was a mixed design study which explored the impact of enactment on working memory for instructions in children with an autism spectrum disorder (ASD), children with an intellectual disability (ID) and typically developing children (TD). The children, aged between 7 and 15 years old, were asked to hear, imagine enacting and enact instruction sequences and recall them orally. In this study a mixed (between and within subject)design has been used (Flick, 2014).

This study focused on the specific enactment condition examined in the study of Xie et al. (2024), and the hearing condition to which working memory performance in the enactment condition can be compared. By using this methodology, the influence and effectiveness of enactment on working memory for instructions could be compared for different groups of children. This study used 2 existing groups, a group with children diagnosed with ASD and a group with TD children. Children in both groups were classified based on characteristics as diagnosis and individual development, classified based on the DSM-V (American Psychiatric Association, 2013).

Sampling

The participants were selected based on specific criteria and had to meet 3 conditions to participate in this study. First, they had to be aged between 7 and 15 years old. Second, they must understand and perform all tasks used in the study. Third, they must be diagnosed with ASD or could be assigned to the TD group. Children who were clinically diagnosed with ASD based on established criteria according to the DSM-V (American Psychiatric Association, 2013) by specialized doctors were included. ASD children with specific medical- or psychiatric comorbidities were excluded. TD children without a history of neurological or psychiatric diagnosis were included. Based on these conditions, the 2 existing groups of children could be compared to each other. Purposive sampling has been used (Flick, 2014). The ASD children were selected from special education schools in China and TD children were selected from regular schools in the same areas in China.

The independence of the test-taking was guaranteed because each respondent participated in the study only once. In addition, the research participants had no mutual contact during the test-taking sessions.

Participants

In the original study of Xie et al. (2024) there was a sample of groups of 20 children (total N=60): 20 children with an autism spectrum disorder (ASD), 20 children with an intellectual disability (ID) and 20 typically developing (TD) children. Each group consisted of children aged between 7 and 15 years old: 4 females and 16 males. The children were all Chinese and right-handed. This research focused on the data collected from both the ASD- and TD groups. The database consisted of 40 children, 20 ASD children and 20 TD children.

The target population are ASD children and TD children aged between 7 and 15 years old. The accessible population are ASD children and TD children aged between 7 and 15 years old from educational institutions in China.

Instruments

First, all children were grouped based on their intelligence to ensure they were comparable in terms of cognitive abilities. To measure the fluid intelligence, the Raven Standard Progressive Matrices Test (SPM; Raven, Raven, & Court, 1998) was used. This non-verbal test consists of 60 figures with missing parts where children had to choose the right answer between several given options to get a point. Based on the total score compared to norm tables from China a level of fluid intelligence was assigned, with SPM results divided from superior intelligence (95%) till mental deficiency (1% and 5% levels) (Zhang & Wang, 1989). A higher score means a child is better at problem-solving and abstract reasoning. The SPM Test and the quality of the manual and testmaterials are rated as good by the COTAN (Evers & Commissie Testaangelegenheden Nederland, 2002). The reliability, construct validity and criterion validity were found to be sufficient. Norms were rated as inadequate due to being outdated, according to the COTAN (Evers & Commissie Testaangelegenheden Nederland, 2002). Raven (2000) points out that cultural differences can affect testresults and interpretation of intelligence in different contexts.

The Peabody Picture Vocabulary Test Revised Chinese Version (PPVT-R; Sang & Miao, 1990) was used to measure crystallized intelligence. This test consists of 180 items where children must choose the right picture that suits the best after hearing a word to get a point. The total score was compared to norms tables from China and intelligence is not divided into levels. Verbal skills could be measured, a higher score means a child has a larger receptive vocabulary. All children had to meet the criteria where they have a similar low-, medium- or high level of fluid and crystallized intelligence. Both ASD- and TD children had SPM scores between 50 and 70%, which made these groups comparable. The construct validity and norms of the PPVT-R are rated as sufficient by the COTAN (Evers & Commissie Testaangelegenheden Nederland, 2002). However, it should be taken into account that the norms are more than 20 years old and might be outdated. The reliability, despite missing test-retest data for children, is rated as good just like the quality of the manual and testmaterials (Egberink & Leng, 2009-2024). There is no clarity about the validity (Dunn & Dunn, 2005), because not enough research was conducted on this (Evers & Commissie Testaangelegenheden Nederland, 2002).

Only children in the ASD group were required to take 2 additional tests to confirm the ASD diagnosis and its severity. First, they had to do the Chinese Childhood Autism Rating Scale (CARS-CV; Schopler et al., 1980). This is an observation from a behavioral scientist evaluating the child on 15 behavioral indicators, scored from 1 till 4. In the study of Xie et al. (2024) the cutoff point for ASD was 30. The CARS Test is considered as a reliable instrument, with a high intern consistency (α=.94) (Rellini et al., 2004) and a high interrater reliability (r=.71) (Schopler et al., 1980). Second, the Autism Behavior Checklist (ABC-CV; Yang et al., 1993) was used. 57 behavioral items about the child were filled in by an assessor, scaled from 0, absent behavior, till 3, behavior is seen often. In the study of Xie et al. (2024) the cutoff point for ASD was 31. In both the CARS and ABC tests a higher score means a child has more autistic behaviors. Literature states important issues of reliability and validity of the ABC remain to be addressed (Rellini et al., 2004). The validity can vary across different subscales, with correlations from insufficient (r=.37) to good (r=.92) (Krug, Arick & Almond, 1988). The reliability of the overall ABC test is rated as good (α=.89), but Volkmar et al. (1988) found a split-half reliability of .70, indicating the subscales are not sufficiently reliable. The interrater reliabilities were strong (κ>.75) for only a minority of items (Volkmar et al., 1988). Therefore, it is important and recommended to combine this instrument with more specific and valid diagnostic tools (Rellini et al., 2004), as done with the CARS Test in this study.

The Span task (Daneman & Carpenter, 1980) has been used for measuring working memory capacity for enactment in instructional tasks. In total, 14 objects were subjected to 6 movements to develop action phrases for the children. The objects used in the task were scissors, a key, a pencil, a comb, a towel, a cup, a tissue, an umbrella, a spoon, a glove, chopsticks, a mirror, a mask and a toothbrush. The movements spin, touch, push, shake, knock and turn over were used (Xie et al., 2024). The participants had 3 seconds to physically enact the instruction with their hands after each spoken instruction. When 3 out of 4 memorized sequences were correct, the next span was started with a maximum span of 5. Scoring is based on the number of correct sequences. Each span is worth 1 point. The span score of enactment has been calculated with the formula $N + n * 0.33$, with N as the number of passed spans and n as the number of correct sequences in the failed span. The test stopped when 2 or less of the 4 sequences could be recalled (Xie et al., 2024).

Procedures

The parents or caretakers of the participants provided written consent for participating in the study after receiving information about the study. Terms of participation, how the collected data will be used and the fact that participants can withdraw from the study without consequences at any given point were described in the informed consent (Xie et al., 2024).

Before starting the main experiment, the participants were familiarized with naming all objects and manipulating these as instructed. In this way the children were enabled to be comfortable with using the materials. Each participant had an individual session with the researcher in a quiet place at school. A trained researcher performed the instruction sequences orally on-site for the Span task (Xie et al., 2024). Participants had to manipulate objects between action phrases. They had to recall the sequences in the correct order presented by the researcher within 1 minute, with 2 practice trials before each condition. In total the experiment took 20 to 30 minutes per participant. The exact period of data-collection was not stated in the original study of Xie et al. (2024).

Data analysis

To perform the data analysis, the statistical software program JASP (version 0.18.3.0;JASP Team, 2018) has been used. First the group ‘ID’ (intellectually disabled) children have been filtered out by using the ‘show’ filter in JASP. Then the ‘enactment’ and ‘hearing’ span score variables were changed from a nominal variable (which they default to in JASP) to a scale variable in the dataset so they can be used as continuous dependent variables. After that, the assumptions of the Repeated Measures ANOVA and -ANCOVA were tested. The assumption of normality was not met for both tests and the assumption of linearity between the covariate and dependent variables was not met for the Repeated Measures ANCOVA. When data is not normally distributed, the correlation between two variables may be a poor measure of association (Kowalski, 1972). Inaccurate p-values produced by the F-test can lead to problems with the statistical significance. The more distorted the distribution is, the more likely it is to commit either a Type I error or a Type II error (Courtney & Chang, 2018). In this study a small sample size was used (N=40), allowing increased variability and sensitivity to extreme values to have a disproportionately large effect on outcomes, which can result in non-normality. In smaller sample sizes the Central Limit Theorem is less effective, because the sample mean may still exhibit significant skewness (Courtney & Chang, 2018). Skewed data may result in losing its ability to detect true effects, a reduction of power (Siebert & Siebert, 2017). Therefore, it was chosen to do a log-transformation to normalize the data by minimizing potential distortion and bringing the values into a more manageable range (Changyong et al., 2014). The log-transformation was performed on hearing (LogHearing), enactment (LogEnactment) and age (LogAge). However, this still might not be enough to ensure the validity of the tests because transformations sometimes only partially improve the distribution or satisfy assumptions (Field, 2013).

After checking the assumptions, a 2x2 Repeated Measures ANOVA was performed to gather information about the enactment effect on working memory. It compares the means of different groups and determines whether there is a statistically significant difference (Field, 2013). In the case of this study, ANOVA helps provide insight into how enactment helps with working memory capacity compared with a baseline of hearing. Hearing, enactment, ASD- and TD children are the conditions of interest. Hearing and enactment are two categories in the within-subjects variable. Developmental group, comprised of ASD- and TD children, is the independent variable. Differences in averages between both ASD- and TD groups on the Span task were analyzed, checking if TD children perform significantly better than ASD children on working memory capacity for instruction tasks. The influence of age on the relationship between enactment and working memory was measured using the Repeated Measures ANCOVA. Performing this test could provide insight into whether age is a predictor for the influence of enactment on working memory for both ASD- and TD children. Possible age-related growth patterns could thereby be identified.

In the original study of Xie et al. (2024) it is not stated if there were missing data, how the missing data was handled, and why it was missing. To avoid bias, multiple imputation would likely have been the best strategy to manage missing data in the case that the missing data were not missing at random (Carpenter et al., 2023).

To measure the effect size of the ANOVA and ANCOVA, η^2p was used and interpreted in the same way for both models. The η^2p represents the proportion of variance explained by a specific factor, adjusted for other effects and influences (Pierce, Block & Aguinis, 2004). The η^2p was chosen because it offers a clearer view off a given effect's strength, and it corrects for noise of error variance for each model (Norouzian, 2018).

Validity and reliability

The children in both groups were categorized based on their fluid and crystallized intelligence levels (Xie et al, 2024). This made them comparable in terms of cognitive abilities, which helps strengthen the validity and reliability of the study. Potential differences in cognitive function that might confound the results could therefore be minimized. Differences in working memory can be more confidently attributed to the enactment condition, rather than to the children's overall cognitive abilities. Additionally, standardized procedures were used to measure the influence of enactment on working memory in both groups. Consistent application of predetermined methods reduces the likelihood of systematic error and the influence of external factors (Maxwell, Delaney & Kelley, 2017). This makes this study's findings more consistent and replicable.

In this study, multiple tests were used to answer the research questions. Because there was no information available on the evaluation and validity and reliability of the Chinese versions, Dutch and English evaluations were performed. In the original study of Xie et al. (2024), Chinese versions of tests were used. Using tests created for a specific population from which the sample was drawn usually increases the validity for this specific context (Flick, 2014). However, because of the fact there is no accessible information on the Chinese versions, there is some uncertainty about the psychometrics properties of it. Therefore, statements whether these instruments are valid and reliable cannot be made. Based on the other samples they have been verified in, it is expected that it should perform similarly with Chinese children and the population that has been used in this study.

Results

In this study, a p-value of <0.05 is considered statistically significant, and therefore, in the context of this study, a meaningful measure of the detection of an experimental effect.

Repeated Measures ANOVA

The 2x2 Repeated Measures ANOVA was used to determine if there is a statistical difference between the means of the two independent groups. To draw good conclusions, it is important to test statistical assumptions when doing research (Verma & Abdel-Salam, 2019). For a detailed description of the assumptions, see Appendix A. Because the assumption of normality was not met, these results should be interpreted with caution (in this case, because the sample is small, the ANOVA is not very robust to violations of normality, meaning the results might contain inaccurate p-values or reduce statistical power).

Table 1 shows a significant effect of the instruction type on the independent variable, the working memory performance ($p < .001$). There is a large effect of the instruction type in working memory performance as measured by the span score ($\eta^2 p = .835$). This means 83.5% of this variation in the data is explained by the instruction type. The results also show a significant interaction effect of instruction and group on the dependent variable, the working memory performance ($p = .044$). A small effect has been found ($\eta^2 p = .077$). The effect of instruction, using enactment or traditional verbal instruction, differs between ASD- and TD children with a benefit from instruction type for TD children.

Table 1

Within-Subjects Repeated Measures ANOVA for instruction and group effects on working memory performance

Cases	Sum of Squares	df	Mean Square	F	P	η^2_p
Instruction	1.643	1	1.643	192.524	< .001	0.835
Instruction ✻ group	0.027	1	0.027	3.159	0.044	0.077
Residuals	0.324	38	0.009			

Note. Type III Sum of Squares

Table 2 shows a significant effect of group, ASD- and TD children (p=.004). A small effect has been found ($\eta^2 p$=.202). There is a difference in the working memory performance between ASD- and TD children, regardless of the instruction type.

Table 2

Between-Subjects Repeated Measures ANOVA for group effects on working memory performance

Cases	Sum of Squares	df	Mean Square	F	P	η^2_p
Group	0.253	1	0.253	9.595	0.004	0.202
Residuals	1.002	38	0.026			

Note. Type III Sum of Squares

Repeated Measures ANCOVA

The Repeated Measures ANCOVA was used to determine the influence of age on the relationship between enactment and working memory capacity. Possible age-related growth patterns could thereby be identified. Before performing the Repeated Measures ANCOVA, its assumptions were checked (Appendix B). Because the assumptions of normality and a linear relationship between the covariate and dependent variables were not met, the results should be interpreted with caution.

Table 3 shows the effect of instruction on working memory performance is not significant (p =.269). There is a small effect size ($\eta^2 p$=.033). This means there is no difference in the working memory performance for the different instruction types, after adjusting the covariate age. There is a significant effect in the interaction between the introduction type and group, ASD- and TD children

(p=.034). A large effect has been found (η^2p=.078). Therefore, it can be said the relationship between instruction type and the working memory performance differs between both groups. The influence of enactment is different for ASD children than for TD children. There is no significant effect of the interaction between instruction and age (p=.661). This means that age does not significantly moderate the effect of instruction type on the working memory performance.

Table 3

Within-Subjects Repeated Measures ANCOVA for instruction and interaction effects on working memory performance

Cases	Sum of Squares	df	Mean Square	F	p	η^2_p
Instruction	0.011	1	0.011	1.258	0.269	0.033
Instruction ✻ group	0.027	1	0.027	3.150	0.034	0.078
Instruction ✻ age	0.002	1	0.002	0.195	0.661	0.005
Residuals	0.323	37	0.009			

Note. Type III Sum of Squares

Table 4 shows a significant effect of group (p=.004). There is a large effect size (η^2_p=.201). This means significant differences have been found for working memory performance between both groups, independent of the instruction type. There is no significant effect of age as a covariate (p=.966). There is a trivial effect size (η^2p=.00007), which means the proportion of the variance explained by age is extremely low. This indicates there is no significant influence of age on the working memory performance, including group differences and instruction type.

Table 4

Between-Subjects Repeated Measures ANCOVA for group and age effects on working memory performance

Cases	Sum of Squares	df	Mean Square	F	p	η^2_p
Group	0.252	1	0.252	9.317	0.004	0.201
Age	4.889×10^{-5}	1	4.889×10^{-5}	0.002	0.966	6.963×10^{-5}
Residuals	1.002	37	0.027			

Note. Type III Sum of Squares

Conclusion and discussion

This study focused on gaining insight into the influence of enactment on working memory of children with an autism spectrum disorder (ASD) and typically developing children (TD) and the role of age on this relation. First, this study attempted to answer the research question: *Does the action-based processing type enactment improve working memory capacity for instructions differently for children on the autism spectrum (ASD) and typically developing children (TD), compared to traditional verbal instructions?* In this study a significant effect was found in both group and instruction type on the working memory capacity of ASD- and TD children. This means there is an effect of enactment on working memory capacity regardless of the group, with a benefit for TD children. TD children therefore might experience that enactment is more effective for their working memory capacity than for ASD children.

In line with previous research about the influence of enactment on working memory (Wang et al., 2022; Engelkamp, 1998) a positive effect for both ASD- and TD children has been found. Despite the fact ASD children show a smaller effect than TD children, implementing enactment tasks in education can improve their overall cognitive function, including working memory (Summers & Craik, 1994). However, as described in the study of Xie et al. (2024), there is, generally, an impaired function of the central executive system in children with ASD. Their episodic buffer shows a reduced ability to integrate different types of information (Lecouvey et al., 2015). This means they may struggle in their daily life with managing and organize their thinking, like following multiple-step instructions and tasks or understanding situations where a lot of information is given. They may also experience challenges in understanding both implicit social and cognitive cues (Minshew & Williams, 2007). Enactment is an action-based instruction type that requires combining multiple types of information at once (Wang et al., 2024). ASD children often show atypical patterns of attention and visual processing, which diminishes the ability to integrate actions with verbal instructions (Pelphrey, Morris & McCarthy, 2005). This could be an explanation for the fact enactment is less effective for ASD children than TD children.

In recent years, there has been a shift from traditional teaching methods and instruction to more interactive and student-centered approaches (Darling-Hammond, Hyler & Gardner, 2017). Literature shows educators are actively seeking effective strategies for supporting ASD children in mainstream classrooms (Dukpa et al., 2021). Because enactment appears to have a positive effect on working memory performance, it might help children with their achievements in the classroom. Dawson et al. (2010) state that using enactment makes following multi-step instructions and switching between tasks easier. Because the effect of enactment appears to be smaller for ASD children, it is therefore questionable whether using only enactment is effective enough for them.

Processing auditory information is more difficult for ASD children than processing visual information (Knight, Sartini & Spriggs, 2015). ASD children often benefit from structured, visual or tangible learning approaches (Tager-Flusberg & Kasari, 2013). It may be interesting to integrate instructional techniques that combine these factors with enactment. Implementing this in the classroom can, for example, be done by using pictograms or creating a visual timeline with certain actions or tasks that children must physically carry out. The study of Fiorini (2017) is an effective example of this in practice. The activity routines of ASD children improved when they had to actively retell and illustrate the teacher's speech combined with using pictograms as well.

Second, the research question: *What is the influence of age on the relationship between the action-based processing type enactment and working memory capacity for instructions in children on the autism spectrum (ASD) and typically developing children (TD)?* had been examined. Results in this study show age appears to have an extremely minimal impact on the working memory performance, whereas group and interaction between instruction type and group significantly influence this.

However, the findings in this study are not consistent with literature about working memory, enactment and the influence of age. Because children are more able to retain and process information when they grow older (Diamond, 2013) and their working memory capacity is similar to adults at the age of 11 years (Gathercole, 1998), better working memory performances were expected for older children. The study of Waterman et al. (2017) about the effect of enactment for instructions in working memory capacity does meet this expectation. They found that the effect of enactment for instructions may increase with age, as older children performed better in processing and responding to instructions involving enactment. The strength of this effect does not apply to this study. This may be due to differences in the sample characteristics in both studies. The study of Waterman et al. (2017) had a larger sample, a more equal gender bias and was taken in a different continent. According to the Central Limit Theorem, the sample mean in small samples may not estimate the population mean and therefore poorly estimates the distribution of the population (Courtney & Chang, 2018). This might be the case with found effect of the age variable in this study. Additionally, the sample of Waterman et al. (2017) consisted of a different age range, measuring the working memory at an earlier point in the lifespan, namely from 6,5 to 10,5 years old, with a mean age of 8 years and 4 months. Between the ages of 7 and 12 years old, there is a rapid growth in the development of working memory (Diamond, 2013). In this study there was an age range in the sample from 7 to 15 years old. Literature states a working memory level equivalent to that of an adult is reached from the age of 11 years old (Gathercole, 1998). This indicates there might be changes in the role of age in working memory over the lifespan, which can provide different outcomes when comparing different age groups.

Additionally, the effect of age may have been influenced by other variables that may have been stronger determinants of working memory performance. Children with neurodevelopmental disorders like ASD show different developmental patterns compared to TD children (Alloway & Gathercole, 2006). Atypical or delayed cognitive development may be associated with slower improvements in domains like working memory and executive functioning (Joseph et al., 2005). One factor that might explain why age had no significant effect on working memory performance, is that there could be a difference between the developmental curves of ASD- and TD children, in which there is a discrepancy between the growth in working memory and age in both groups. It is also possible that the Span task was not adequately adapted to the children's developmental levels. Conducting tasks that are too complex can lead to limited performance for both younger and older children as well (Diamond, 2013).

Overall, it is remarkably interesting to look at the value of enactment on working memory for other children than ASD- and TD children as well. For example, children with learning disabilities (Gathercole & Alloway, 2008) or ADHD (Craig et al., 2016) can experience problems with their executive functions such as working memory as well. The physical aspect of enactment can also support children with motor problems, because of the interaction between motor and cognitive processes (Schneider & Pressley, 2013). It is therefore likely that implementing enactment can contribute to inclusive education, when meeting the needs of multiple children (Lynch & Irvine, 2009).

Strengths and limitations

Several factors may have contributed to the fact that the results found in this study do not fully correspond with findings in the literature. The data should therefore be interpreted with caution.

Data collection and sample size

The selection of research participants was not done randomly, but purposive sampling has been used (Flick, 2014). In the original study of Xie et al. (2024) test-taking has therefore taken place in different schools across China. According to Robson (2016), measuring moments must be as similar as possible to compare them correctly and obtain reliable data. When using standardized procedures in test-taking, like assessing all participants under similar conditions, systemic error and external influences could be minimized (Maxwell, Delaney & Kelley, 2017). This increases the research validity and ensures any differences in performance are not due to the procedure. However, it should be considered that the test environment, location and school, differed for the children in this study. This might be related to the individuals' attention and concentration, which could have affected their working memory performance as well.

Besides that, in the original study of Xie et al. (2024) data from only 20 children of both ASD- and TD groups was collected. Due to this small sample size, the results must be interpreted with consideration. A small sample size reduces statistical power with an increasing risk of sampling error, limiting its validity (Field, 2013). Additionally, the smaller a study sample is, the less likely it is to adequately capture population characteristics (Courtney & Chang, 2018).

Missing data and Assumptions

Remarkably, the dataset of the study of Xie et al. (2024) was a complete dataset and did not contain any missing data. The fact there is no transparent reporting of handling missing data – even just to simply confirm that no data were missing – is concerning, because it is crucial for the reliability and validity of the findings (Enders, 2022). This possibly indicates an unrealistic scenario of there being no missing data (human participant studies tend to encounter attrition, for instance, especially when the participants are children) or handling of missing data without declaring how the data were handled and why, creating a biased view of the population (Little & Rubin, 2019).

Additionally, the period of data collection was not mentioned which leads to implications on different levels. Factors such as time of sampling, as well as time of test-taking may affect outcomes and are essential for the reproducibility and validity of the research (Little & Rubin, 2019). For example, if the data collection took place during or immediately after the COVID pandemic, the results may be biased by the social and psychological factors associated with experiencing a pandemic. This pandemic resulted in significant learning delays and reduced student achievement, for instance. Distance learning has also been influential on the quality of the teacher instructions (Zambach & Hansen, 2023). Multiple factors were affecting both learning and teaching performance (Kuhfeld et al., 2022), and if the tests were taken during this period this may adversely affect the outcomes.

A third point to mention relates to the ANOVA assumptions mentioned earlier. As normality of the residuals was not achieved for some of the conditions, there is a good chance that the results of the ANOVA are not reliable. In some studies that have large samples, ANOVA is somewhat resistant to some violations of normality, but in this case we have a small sample. This means that we risk either Type I error (finding significance erroneously) or Type II error (low power to detect an effect). The implication of this is that the findings of this study might not be totally accurate (at best), or unreliable or wrong (at worst).

Tests

Another factor that may contribute to not meeting the expectations based on the literature is the fact relatively old tests were used, some specifically for the Chinese population. According to the COTAN, a Dutch committee that evaluates the quality of psychological and educational tests, the criterion validity of the SPM is insufficient and the criterion validity of the PPVT was not assessed for example (Egberink & Leng, 2009-2024). This limits opportunities for generalizing the results to other cultures and contexts, where these tests may not function similarly in the Chinese or Dutch population for example.

Using outdated tests can be at the expense of the validity of a study and can lead to inaccurate and misleading results (Hopster-den Otter et al., 2019). Norms tables in these tests are based on old samples and have been used in a context other than that for which they were designed, which may not be representative of the current population. It is questionable whether these tests are relevant for modern children because norms are sensitive to demographic, cultural and educational changes (delCacho-Tena et al., 2024). Dutch literature states that updating norms tables is crucial to ensure validity and reliability of psychological tests (Nederlands Instituut van Psychologen, 2023).

Gender

In the dataset of Xie et al., (2024) there is a gross imbalance between the number of boys and girls (only 12 girls of the total sample 60 participants were involved in this study). Based on the literature used in this research, it is plausible that gender plays a role in the influence of enactment for working memory capacity. Evidence for gender differences on brain structure and function has been found in both normal brain development (Rubia et al., 2010) and individuals with ASD (Gur et al., 2002). The research of Rubia et al. (2010) states activation of brain regions, which crucial in working memory, varies by gender. Men perform better on working memory tasks involving selective attention and shifting because their inferior parietal lobe is activated more than in women. In contrast, research by Huang (1993) shows girls in a Chinese population scored significantly higher on specific working memory tasks than boys. Structural brain differences, like abnormalities in the cerebral cortex and neural connectivity, are responsible for both social and sensory processing in individuals with ASD (Gur et al., 2002). Therefore, the influence of enactment may manifest differently for boys and girls in both TD and ASD individuals, which is why these groups should be imaged separately in follow-up research. Unlikely the data of Xie et al. (2024) represents boys better than girls, giving the gender ratio. There is a possibility that there are interesting gender effects about which no valid statements can be made based on the current research, and the influence of enactment on the working memory performance of girls appears to remain underexposed. Regardless of the true effect gender might have, it is plausible that it confounded the findings based on these data and requires that the results be viewed with caution.

Cultural approach

The tests in the study of Xie et al. (2024) were conducted to a Chinese population. This causes the generalization of the findings to be critically assessed (Flick, 2014). Relying on a sample from one culture or country leads to potential cultural bias and limited applicability of the findings (Matsumoto & Hwang, 2013). It is therefore difficult to say whether we can expect these findings to apply in other countries as well, but it is a reasonable assumption. In line with this study, the study of Macedonia, Müller & Friederici (2011) on the effect of working memory found that physically enacting benefits memory performance as well in another context, with German children.

Cultural differences in psychological research can cause questions and test items to be interpreted differently in several countries, which does not improve reliability and validity (Smith & Bond, 2019). To make the findings more generalizable, it is important to know whether the instruments measure the same construct in different cultures. This will ensure that the results are valid in other countries and contexts (van de Vijver, 2015).

Recommendations

Based on the limitations of this study, there are a number of recommendations for further research. To make this study more generalizable, a bigger study needs to be done with a broader sample of children, using a more equal division between boys and girls. A larger sample size enhances the likelihood the research conclusion can be applied to the broader population (Button et al., 2013). To prevent gender-bias, it is important to have an equal distribution of gender in the sample (Heidari et al., 2016). Including multiple cultures and nationalities in a sample prevents cultural bias and therefore can increase the generalizability of the study (Greenfield, 2017). Describing the data collection and how missing data was handled more specifically can also contribute to the generalizability (Field, 2013), and at the very least, declarations of missing data and the handling approach can ensure that people can use the dataset in the future without concerns that bias has been introduced into the data set through improper missing data handling procedures (for instance, if the data were not missing at random, but the missing cells were handled by case-wise deletion).

The data in this study was collected at one specific moment. It is recommended to conduct a longitudinal study of the influence of enactment on the working memory of ASD- and TD children. In fact, longitudinal research may be of specific interest to ASD children because they might develop differently from normally developing children (Landa, Holman & Garett-Mayer, 2007). Insights into correlations and causal relationships can be examined by longitudinal research (Twisk, 2023), therefore it might be interesting to test for variations in teachers' instructions. Changes in working memory performance, the long-term effects of enactment and whether it remains effective can thereby be monitored.

The results in this study show that enactment may be less effective for ASD children than for TD children, should the p-values be somewhat accurate. It is therefore interesting in future research to look at the individual differences within the group with ASD children to understand what makes enactment less effective for them, especially with a larger sample. There could be a focus on severity of ASD, language- and social skills. Research of Tager-Flushberg & Kasari (2013) shows for example that children with more severe levels of autism have more impairments that negatively affect academic achievement, including listening to instructions. Further research could potentially lead to the development and implementation of personalized instructional strategies for this target population.

Implications

This study contributed to understanding the effect of enactment on working memory performance for ASD- and TD children, and the role of age on this relationship. Because a significant effect of enactment on working memory capacity for both ASD- and TD children has been found, it is interesting to do further research on this.

This research also offers valuable insights in practice for education and teachers. Because of the found effect that enactment can improve children's working memory performance, it can be a successful strategy in the classroom. A child's learning environment and academic performance is influenced by their teachers' actions and instructions (Darling-Hammond, Hyler & Gardner, 2017). On the one hand, which is why is important to raise awareness among teachers, but also professionals and other stakeholders in education, about the importance of using enactment the classroom. There are different ways to educate teachers about implementing enactment in their teaching style. Active participation by attending workshops or courses, perceiving feedback by instructional coaches or collaborating in learning communities where teachers can share experiences are ways in which knowledge can be contributed to them (Darling-Hammond, Hyler & Gardner, 2017).

On the other hand, there appears to be a need from teachers to implement strategies for inclusive education (Petersson-Bloom & Holmqvist, 2022). A teacher's disposition and behavior are seen as key factors to successful inclusion (Burack, Root, & Zigler, 1997). Therefore, it is important to encourage enactment in the instructional tasks. When meeting the needs of more children, differences within the classroom can be reduced (Lynch & Irvine, 2009). This aligns with the vision expressed in the Salamanca Statement (UNESCO, 1994), that children have the right and should be able to join general schools.

References

Alloway, T. P. (2006). How does working memory work in the classroom?. *Educational Research and reviews,* 1(4), 134-139.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders:* DSM 5. American Psychiatric Publishing. American Psychiatric Association. (2013).

American Psychiatric Association. (2022). What is autism spectrum disorder? Psychiatry.org. https://www.psychiatry.org/patients-families/autism/what-is-autism-spectrum-disorder

Ashburner, J., Ziviani, J., & Rodger, S. (2008). Sensory processing and classroom emotional, behavioral, and educational outcomes in children with autism spectrum disorder. *The American Journal of Occupational Therapy, 62*(5), 564-573.

Baddeley, A. (2010). Working memory. *Current biology*, *20*(4), R136-R140.

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual review of psychology*, *63*(1), 1-29.

Baddeley, A. D. (2003a). Working memory and language: An overview. *Journal of Communication Disorders,* 36, 189-208.

Bolourian, Y., Losh, A., Hamsho, N., Eisenhower, A., & Blacher, J. (2021). General education teachers' perceptions of autism, inclusive practices, and relationship building strategies. *Journal of autism and developmental disorders*, 1-14.

Brockmole, J. R., & Logie, R. H. (2013). Age-related change in visual working memory: A study of 55753 participants aged 8-75. *Frontiers in Psychology*, 4, 12.

Burack, J. A., Root, R., & Zigler, E. (1997). Inclusive education for students with autism: Reviewing ideological, empirical, and community considerations. *Handbook of autism and pervasive developmental disorders, 2*, 796-807.

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature reviews neuroscience*, *14*(5), 365-376.

Carpenter, J. R., Bartlett, J. W., Morris, T. P., Wood, A. M., Quartagno, M., & Kenward, M. G. (2023). *Multiple imputation and its application*. John Wiley & Sons.

Changyong, F. E. N. G., Hongyue, W. A. N. G., Naiji, L. U., Tian, C. H. E. N., Hua, H. E., & Ying, L. U. (2014). Log-transformation and its implications for data analysis. *Shanghai archives of psychiatry*, *26*(2), 105.

Courtney, M. G., & Chang, K. C. (2018). Dealing with non-normality: an introduction and step-by-step guide using R. *Teaching Statistics*, *40*(2), 51-59.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of verbal learning and verbal behavior*, 19(4), 450-466.

Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). Effective teacher professional development. *Learning policy institute*.

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.

delCacho-Tena, A., Christ, B. R., Arango-Lasprilla, J. C., Perrin, P. B., Rivera, D., Olabarrieta-Landa, L. (2024). Normative Data Estimation in Neuropsychological Tests: A Systematic Review. *Archives of Clinical Neuropsychology*, 39(3), 383–398.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, 64, 135–168. https://doi.org/10.1146/annurev-psych-113011-143750.

Dukpa, D., Carrington, S., & Mavropoulou, S. (2021). Exploring Bhutanese teachers' knowledge and use of strategies for the inclusion of students on the autism spectrum. *International Journal Of Inclusive Education*, 1–24. https://doi.org/10.1080/13603116.2021.1973124

Dunn, L. M., Dunn, L. M., & Schlichting, L. (2005). *Peabody Picture Vocabulary Test-III-NL*. Lisse: Harcourt Test Publishers.

Egberink, I.J.L. & Leng, W.E. de. (2009-2024). COTAN Documentatie (www.cotandocumentatie.nl). Amsterdam: Boom Uitgevers Amsterdam.

Enders, C. K. (2022). *Applied missing data analysis*. Guilford Publications.

Engelkamp, J. (1998). *Memory for actions.* Psychology Press.

Eric, F. (2020). Epidemiological controversies in autism. *Swiss Archiv Neurol, Psychiatr Psychotherapy*, *171*(1), 3-5.

Evers, A., & Commissie Testaangelegenheden Nederland. (2002). *COTAN testboek voor het onderwijs*. NDC ; Boom.

Field, A. (2013). Discovering statistics using IBM SPSS statistics.

Fiorini, B. S. (2017). Students with autism spectrum disorders in early childhood education: characterization of the school routine.

Flick, U. (2014). *An introduction to qualitative research* (5th ed.). Sage Publications Ltd.

Freitag, S., & Dunsmuir, S. (2015). The inclusion of children with ASD: Using the theory of planned behaviour as a theoretical framework to explore peer attitudes. *International Journal of Disability, Development and Education*, *62*(4), 405-421.

Gathercole, S. E. (1998). The Development of Memory. *Journal Of Child Psychology and Psychiatry*, *39*(1), 3–27. https://doi.org/10.1017/s0021963097001753

Gathercole, S. E. (2004). Working memory and learning during the school years. In *Proceedings of the British Academy* (Vol. 125, No. 1, pp. 365-380).

Gathercole, S. E., Lamont, E. M. I. L. Y., & Alloway, T. P. (2006). Working memory in the classroom. *In Working memory and education* (pp. 219-240). Academic Press.

Craig, F., Margari, F., Legrottaglie, A. R., Palumbi, R., De Giambattista, C., & Margari, L. (2016). A review of executive function deficits in autism spectrum disorder and attention-deficit/hyperactivity disorder. *Neuropsychiatric disease and treatment*, 1191-1202.

Grainger, C., Williams, D. M., & Lind, S. E. (2014). Online action monitoring and memory for self-performed actions in autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 44(5), 1193–1206. https://doi.org/10.1007/s10803-013-1987-4

Greenfield, P. M. (2017). Cross-cultural research and Piagetian theory: Paradox and progress. In *The developing individual in a changing world* (pp. 322-333). Routledge.

Gur, R. C., Gunning-Dixon, F., Bilker, W. B., & Gur, R. E. (2002). Sex differences in temporo-limbic and frontal brain volumes of healthy adults. *Cerebral cortex*, *12*(9), 998-1003.

Heidari, S., Babor, T. F., De Castro, P., Tort, S., & Curno, M. (2016). Sex and gender equity in research: rationale for the SAGER guidelines and recommended use. *Research integrity and peer review*, *1*, 1-9.

Hopster-den Otter, D., Wools, S., Eggen, T. J., & Veldkamp, B. P. (2019). A general framework for the validation of embedded formative assessment. *Journal of Educational Measurement*, *56*(4), 715-732.

Howell, D. C., Rogier, M., Yzerbyt, V., & Bestgen, Y. (1998). Statistical methods in human sciences. *New York: Wadsworth*, *721*.

Huang, J. (1993). An investigation of gender differences in cognitive abilities among Chinese high school students. *Personality and individual differences*, 15, 717-719.

JASP Team. (2018). *JASP (Version 0.18.3) [Computer software]*.

Joseph, R. M., Steele, S. D., Meyer, E., & Tager-Flusberg, H. (2005). Self-ordered pointing in children with autism: failure to use verbal mediation in the service of working memory?. *Neuropsychologia*, *43*(10), 1400-1411.

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: the contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of experimental psychology: General*, 132(1), 47.

Kern, J. K., Trivedi, M. H., Garver, C. R., Grannemann, B. D., Andrews, A. A., Savla, J. S. & Schroeder, J. L. (2006). The pattern of sensory processing abnormalities in autism. *Autism, 10*(5), 480-494.

Klein, K. & Saltz, E. (1976). Specifying the mechanisms in a levels-of- processing approach to memory. *Journal of Experimental Psychology: Human Learning and Memory*, 2, 671-679.

Knight, V., Sartini, E., & Spriggs, A. D. (2015). Evaluating visual activity schedules as evidence-based practice for individuals with autism spectrum disorders. *Journal of autism and developmental disorders, 45*, 157-178.

Kowalski, C. J. (1972). On the effects of non-normality on the distribution of the sample product-moment correlation coefficient. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 21*(1), 1-12.

Krug, D. A., Arick, J., & Almond, P. (1988). Autism behavior checklist. *Child Psychology & Psychiatry & Allied Disciplines*.

Kucharczyk, S., Reutebuch, C. K., Carter, E. W., Hedges, S., El Zein, F., Fan, H., & Gustafson, J. R. (2015). Addressing the needs of adolescents with autism spectrum disorder: Considerations and complexities for high school interventions. *Exceptional Children, 81*(3), 329-349.

Kuhfeld, M., Soland, J., & Lewis, K. (2022). Test Score Patterns Across Three COVID-19-Impacted School Years. *Educational Researcher, 51*(7), 500-506. https://doi.org/10.3102/0013189X221109178

Landa, R. J., Holman, K. C., & Garrett-Mayer, E. (2007). Social and communication development in toddlers with early and later diagnosis of autism spectrum disorders. *Archives of general psychiatry*, *64*(7), 853-864.

Lecouvey, G., Quinette, P., Kalpouzos, G., Guillery-Girard, B., Bejanin, A., Gonneaud, J. & Desgranges, B. (2015). Binding in working memory and frontal lobe in normal aging: is there any similarity with autism?. *Frontiers in human neuroscience*, *9*, 90.

Little, R. J., & Rubin, D. B. (2019). *Statistical analysis with missing data* (Vol. 793). John Wiley & Sons.

Lobar, S. L. (2016). DSM-V changes for autism spectrum disorder (ASD): Implications for diagnosis, management, and care coordination for children with ASDs. *Journal of Pediatric Health Care, 30(4)*, 359-365.

Lord, C., Elsabbagh, M., Baird, G., & Veenstra-Vanderweele, J. (2018). Autism spectrum disorder. *The lancet*, *392*(10146), 508-520.

Lynch, S. L., & Irvine, A. N. (2009). Inclusive education and best practice for children with autism spectrum disorder: An integrated approach. *International journal of Inclusive education, 13*(8), 845-859.

Macedonia, M., Müller, K., & Friederici, A. D. (2011). The impact of iconic gestures on foreign language word learning and its neural substrate. *Human brain mapping, 32*(6), 982-998.

Matson, J. L., Wilkins, J., Sevin, J. A., Knight, C., Boisjoli, J. A., & Sharp, B. (2009). Reliability and item content of the Baby and Infant Screen for Children with autism Traits (BISCUIT): Parts 1–3. *Research in Autism Spectrum Disorders*, 3(2), 336-344.

Matsumoto, D., & Hwang, H. C. (2013). Assessing cross-cultural competence: A review of available tests. *Journal of Cross-Cultural Psychology, 44*(6), 849–873. https://doi.org/10.1177/0022022113492891

Maxwell, S.E., Delaney, H.D., & Kelley, K. (2017). Designing Experiments and Analyzing Data: A Model Comparison Perspective, Third Edition (3rd ed.). Routledge. https://doi.org/10.4324/9781315642956

Minshew, N. J., & Williams, D. L. (2007). The new neurobiology of autism: cortex, connectivity, and neuronal organization. *Archives of neurology*, *64*(7), 945-950.

Munakata, Y., Morton, B., & O'Reilly, R. C. O. (in press). Developmental and computational approaches to variation in working memory. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory.* Oxford: Oxford University Press.

Nederlands Instituut van Psychologen (NIP). (2023). *COTAN-beoordeling van psychologische testen*. Retrieved from https://www.cotan.nl

Norouzian, R. (2018). *Methodological Reform in Quantitative Second Language Research: Effect Sizes, Bayesian Hypothesis Testing, and Bayesian Estimation of Effect Sizes* (Doctoral dissertation).

Oberauer, K. (2002). Access to information in working memory: exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 411.

Pelphrey, K. A., Morris, J. P., & McCarthy, G. (2005). Neural basis of eye gaze processing deficits in autism. *Brain*, *128*(5), 1038–1048. https://doi.org/10.1093/brain/awh404

Petersson-Bloom, L., & Holmqvist, M. (2022). Strategies in supporting inclusive education for autistic students—A systematic review of qualitative research results. *Autism & Developmental Language Impairments*, 7. https://doi.org/10.1177/23969415221123429

Pierce, C. A., Block, R. A., & Aguinis, H. (2004). Cautionary note on reporting eta-squared values from multifactor ANOVA designs. *Educational and psychological measurement*, *64*(6), 916-924.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Raven manual: Section 3. Standard progressive matrices.* Oxford Psychologists Press.

Raven, J. (2000). The Raven's progressive matrices: change and stability over culture and time. *Cognitive psychology*, *41*(1), 1-48.

Rellini, E., Tortolani, D., Trillo, S., Carbone, S., & Montecchi, F. (2004). Childhood Autism Rating Scale (CARS) and Autism Behavior Checklist (ABC) correspondence and conflicts with DSM-IV criteria in diagnosis of autism. *Journal of autism and developmental disorders*, *34*, 703-708.

Robson, C. (2016). *Real world research: A resource for users of social research methods in applied settings* (4de ed.). Wiley.

Rubia, K., Hyde, Z., Halari, R., Giampietro, V., & Smith, A. (2010). Effects of age and sex on developmental neural networks of visual–spatial attention allocation. *Neuroimage*, *51*(2), 817-827.

Saltz, E. (1988). The role of motoric enactment (M-processing) in memory for words and sentences. In M. Gruneberg, P. Morris, & R. Sykes (Eds.), *Practical aspects of memory: Current research and issues* (Vol. 1, pp. 408-414). London: Wiley.

Sang, B., & Miao, X. (1990). The revision of trail norm of Peabody Picture Vocabulary Test Revised (PPVT-R) in Shanghai proper. *Psychological Science*, 14(5), 20–25.

Schopler, E., Reichler, R. J., DeVellis, R. F., & Daly, K. (1980). Toward objective classification of childhood autism: Childhood Autism Rating Scale (CARS). *Journal of Autism and Developmental Disorders*, 10(1), 91–103. https://doi.org/10.1007/bf02408436

Schneider, W., & Pressley, M. (2013). *Memory development between two and twenty*. Psychology Press.

Siebert, C. F., & Siebert, D. C. (2017). *Data analysis with small samples and non-normal data:*

Nonparametrics and other strategies. Oxford University Press.

Smith, P. B., & Bond, M. H. (2019). Cultures and persons: Characterizing national and other types of cultural difference can also aid our understanding and prediction of individual variability. *Frontiers in Psychology*, *10*, 2689.

Spranger, T., Schatz, T. R., & Knopf, M. (2008). Does action make you faster? A retrieval-based approach to investigating the origins of the enactment effect. *Psychological Science, 19*(11), 1098-1104. https://doi.org/10.1111/j.1467-9450.2008.00675.x

Summers, J., & Craik, F. I. M. (1994). The effects of subject-performed tasks on the memory performance of verbal autistic children. *Journal Of Autism And Developmental Disorders*, *24*(6), 773–783. https://doi.org/10.1007/bf02172285

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2013). *Using multivariate statistics* (Vol. 6, pp. 497-516). Boston, MA: pearson.

Tager-Flusberg, H., & Kasari, C. (2013). Minimally verbal school-aged children with autism spectrum disorder: The neglected end of the spectrum. *Autism Research*, *6*(6), 468-478. https://doi.org/10.1002/aur.1329

Towse, J. N., Hitch, G. J., & Hutton, U. (1998). A reevaluation of working memory capacity in children. *Journal of memory and language*, *39*(2), 195-217.

Tse, A. C., Lee, P. H., Chan, K. K. S., Edgar, V. B., Wilkinson-Smith, A., & Lai, W. H. E. (2019). Examining the impact of physical activity on sleep quality and executive functions in children with autism spectrum disorder: A randomized controlled trial. *Autism*, *23*(7), 1699–1710. https://doi.org/10.1177/1362361318823910

Twisk, J. W. R. (2023). *Causality in observational longitudinal studies*. In *Applied Longitudinal Data Analysis for Medical Science* (pp. 92-115). Cambridge University Press. https://doi.org/10.1017/9781009288002.007

UNESCO. (1994). The Salamanca Statement and Framework for Action on Special Needs Education. In *UNESCO*. https://unesdoc.unesco.org/ark:/48223/pf0000098427

van de Vijver, F. J. R. (2015). Methodological aspects of cross-cultural research. In M. J. Gelfand, C.-Y. Chiu, & Y.-Y. Hong (Eds.), *Handbook of advances in culture and psychology*, Vol. 5, pp. 101–160). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190218966.003.0003

Verma, J. P., & Abdel-Salam, A. S. G. (2019). *Testing statistical assumptions in research*. John Wiley & Sons.

Volkmar, F. R., Cicchetti, D. V., Dykens, E., Sparrow, S. S., Leckman, J. F., & Cohen, D. J. (1988). An evaluation of the Autism Behavior Checklist. *Journal of Autism and Developmental Disorders, 18*(1), 81–97. https://doi.org/10.1007/BF02211820

Wang, L. J., Xie, T. T., Ma, H., Xu, M., & Xie, X. C. (2022). Subject-performed task effect on working memory performance in children with autism spectrum disorder: The role of intelligence. *Autism Research.* https://doi.org/10.1002/aur.2710

Waterman, A. H., Atkinson, A. L., Aslam, S. S., Holmes, J., Jaroslawska, A., & Allen, R. J. (2017). Do actions speak louder than words? Examining children’s ability to follow instructions. *Memory & Cognition*, 45(6), 877–890. https://doi.org/10.3758/s13421-017-0702-7

Xie, T., Ma, H., Wang, L., & Du, Y. (2024). Can enactment and motor imagery improve working memory for instructions in children with autism spectrum disorder and children with intellectual disability?. *Journal of Autism and Developmental Disorders*, 54(1), 131-142.

Yang, X., Huang, Y., Jia, M., & Cheng, S. (1993). Test report on autism behavior checklist. *Chinese Mental Health Journal*, *7*(6),279–280.

Zambach, S., & Hansen, J. U. (2023). Student and teacher performance during COVID-19 lockdown: An investigation of associated features and complex interactions using multiple data sources. *Plos one*, *18*(10), e0291689.

Zhang, H. C., & Wang, X. P. (1989). Standardization research on Raven's standard progressive matrices in China. *Acta Psychologica Sinica, 21(2),* 113–121.

‘

Appendix A. Assumptions of the Repeated Measures ANOVA

Independent observations

The tests were administered individually with no mutual contact possible, in addition each child participated only once in the study. This means this assumption was met.

The independent variables contain at least two categorical related groups

This assumption is controlled using an appropriate research method design. In this study there are 2 independent variables, condition and group. Hearing and enactment are both within-subjects variables for the condition variable, because each child performed the Span task for each of the hearing and enactment conditions, whereas the between-subjects variables are children either on the autism spectrum (ASD) or typically developing children (TD) for the group variable. Because the independent variables (factors) have at least two categorical related groups (levels), this assumption was met.

Normality

To test for the assumptions of normality of the residuals, a Shapiro-Wilk test was performed. Table 5 shows the descriptive statistics and Shapiro Wilk test both the hearing and enactment variable for ASD- and TD children. The Shapiro-Wilks test shows no significant distribution in the hearing condition for TD children (W=.905, p=.051) and a significant distribution for ASD children (W=.869, p=.011). There is a significant distribution in the enactment condition for TD- (W=.888, p=.024) and ASD children (W=.694, p=< .001). This indicates there is no normal distribution of the data for both groups in the enactment condition and for the ASD children in the hearing condition, with a normal distribution of the data for TD children in the hearing condition. This means the normality assumption is quite violated when using these variables.

Table 5

Descriptive Statistics and Shapiro-Wilk Test for Normality of Residuals

	hearing		enactment	
	ASD	TD	ASD	TD
Valid	20	20	20	20
Missing	0	0	0	0
Mean	2.464	2.664	3.149	3.664
Std. Deviation	0.349	0.419	0.333	0.446
Shapiro-Wilk	0.869	0.905	0.694	0.888
P-value of Shapiro-Wilk	0.011	0.051	< .001	0.024
Minimum	2.000	2.000	2.660	3.000
Maximum	3.000	3.660	4.000	4.330

Note. The hearing condition shows no significant deviation from normality for TD children (*p*=.051), and a significant deviation for ASD children (*p*=.011). The enactment condition shows a significant deviation from normality for TD children (*p*=.024) and ASD children (*p*=< .001).

Both hearing and enactment are variables that might introduce non-normality. Therefore, it was chosen to do a log-transformation to normalize the data by minimizing potential distortion and bringing the values into a more manageable range (Changyong et al., 2014). The log-transformation was performed on hearing (LogHearing) and enactment (LogEnactment).

Table 6 shows the descriptive statistics for both the LogHearing and LogEnactment variable for ASD- and TD children. The Shapiro-Wilks test shows no significant distribution in the hearing condition for TD children (*W*=.912, *p*=.068) and a significant distribution for ASD children (*W*=.875, *p*=.014). There is a significant distribution in the enactment condition for TD- (*W*=.877, *p*=.015) and ASD children (*W*=.721, *p*=< .001). This indicates there is no normal distribution of the data for both groups in the enactment condition and for the ASD children in the hearing condition. This means the normality assumption is quite violated as well after doing the log-transformation on both the hearing and enactment condition.

Table 6

Descriptive Statistics and Shapiro-Wilk Test for Normality of Residuals after log-transformation

	LogHearing		LogEnactment	
	ASD	TD	ASD	TD
Valid	20	20	20	20
Missing	0	0	0	0
Mean	0.892	0.968	1.142	1.291
Std. Deviation	0.141	0.157	0.099	0.125
Shapiro-Wilk	0.875	0.912	0.721	0.877
P-value of Shapiro-Wilk	0.014	0.068	< .001	0.015
Minimum	0.693	0.693	0.987	1.009
Maximum	1.009	1.297	1.386	1.466

Note. Note. LogHearing shows no significant deviation from normality for TD children (*p*=.068) and a significant deviation for ASD children (*p*=.014). LogEnactment shows a significant deviation from normality for TD children (*p*=.015) and ASD children (*p*< .001).

When data is not normally distributed, the correlation between two variables may be a poor measure of association (Kowalski, 1972). The p-values produced by the F-test are not accurate, which can lead to problems with the statistical significance. The more distorted the distribution is, the more likely it is to commit either a Type I error or a Type II error (Courtney & Chang, 2018). In this study a small sample size was used (N=40), allowing increased variability and sensitivity to extreme values to have a disproportionately large effect on outcomes, which can result in non-normality. In smaller sample sizes the Central Limit Theorem is less effective, because the sample mean may still exhibit significant skewness (Courtney & Chang, 2018). Skewed data may result in losing its ability to detect true effects, a reduction of power (Siebert & Siebert, 2017). The assumption of normality being violated can affect the findings of the study, which means the results should be interpreted with caution.

Equal variances

To test this assumption, a Levene's test of Equality of Variances was conducted. Table 7 shows that there is no difference in the variance (hearing *p*=.652, enactment *p*=.144). This means the assumption of equal variances is not violated.

Table 7

Test for Equality of Variances (Levene's)

	F	df1	df2	P
hearing	0.207	1	38	0.652
enactment	2.221	1	38	0.144

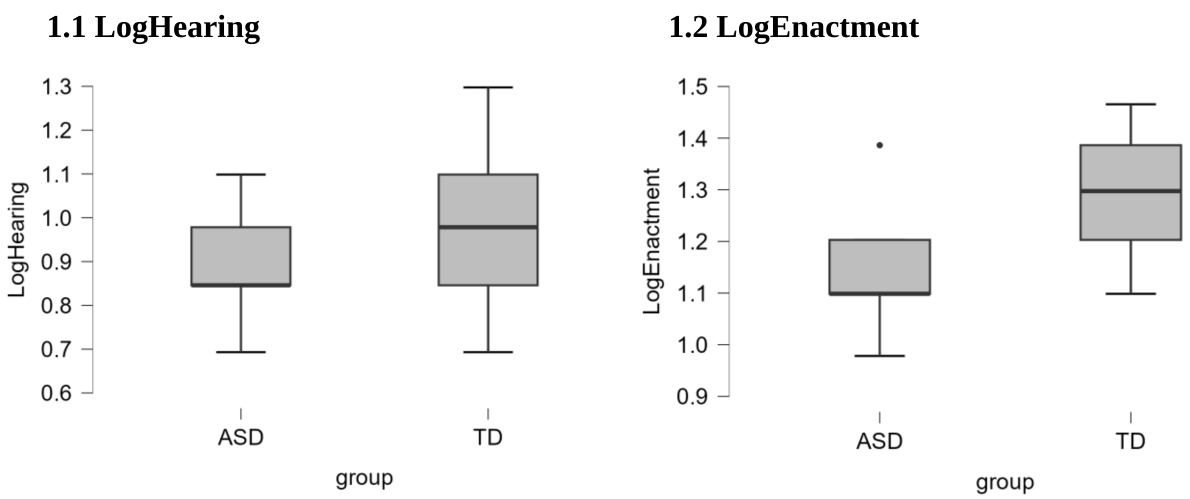
Note. There are no significant differences for the hearing condition (*p*=.652) and the enactment condition (*p*=.144).

Significant outliers

To test for the assumption of significant outliers, a boxplot was made (figure 1). Figure 1 shows there might be an outlier in the enactment condition for ASD children. In a small sample size, outliers could provide essential information about the population. Removing outliers in small samples may bias the results and can reduce the generalizability and representativeness of the study, which is problematic (Tabachnick, Fidell & Ullman, 2013). To detect the presence of an outlier, the method of the mean plus or minus 3 standard deviations can be used (Howell et al., 1998).

Figure 1

Boxplot of scores for LogHearing and LogEnactment in ASD and TD children



Note. The x-axis represents the group (ASD, TD), and the y-axis represents LogHearing (1.1) and Logenactment (1.2). The boxplots demonstrate the distribution of scores for both groups across the conditions, with lower scores for ASD children in both conditions compared to TD children. An outlier is visible in the LogEnactment condition for the ASD group.

Table 7 shows the descriptive statistics of both the LogHearing and LogEnactment for ASD and TD children. Based on the mean (*M*=1.142) and standard deviation (*SD*=.009) in the enactment condition for ASD children, there is a range with an upper value of 1.315 and a lower value of .469. The minimum (.987) and maximum (1.386) are within this range, indicating there is no outlier. This means the assumption of significant outliers was met.

Table 7

Descriptive Statistics for Normality of Residuals after log-transformation

	LogHearing		LogEnactment	
	ASD	TD	ASD	TD
Valid	20	20	20	20
Missing	0	0	0	0
Mean	0.892	0.986	1.142	1.291
Std. Deviation	0.141	0.157	0.099	0.125
Minimum	0.693	0.693	0.987	1.009
Maximum	1.009	1.297	1.386	1.466

Note. This table illustrates the descriptive statistics for ASD and TD children.

Appendix B. Assumptions of the Repeated Measures ANCOVA

Independent observations

The tests were administered individually with no mutual contact possible, in addition each child participated only once in the study. This means this assumption was met.

Linear relationship between the covariate and the dependent variable

The relationship between the covariate and the dependent variable has been examined. Table 8 shows there is a significant relationship for hearing and enactment relationship (r = .551, p < .001). There is no significant relationship for both enactment and age (r=-0.009, p =0.958) and hearing and age (r=.029, p =.857).

Table 8

Pearson's Correlations for hearing, enactment and age

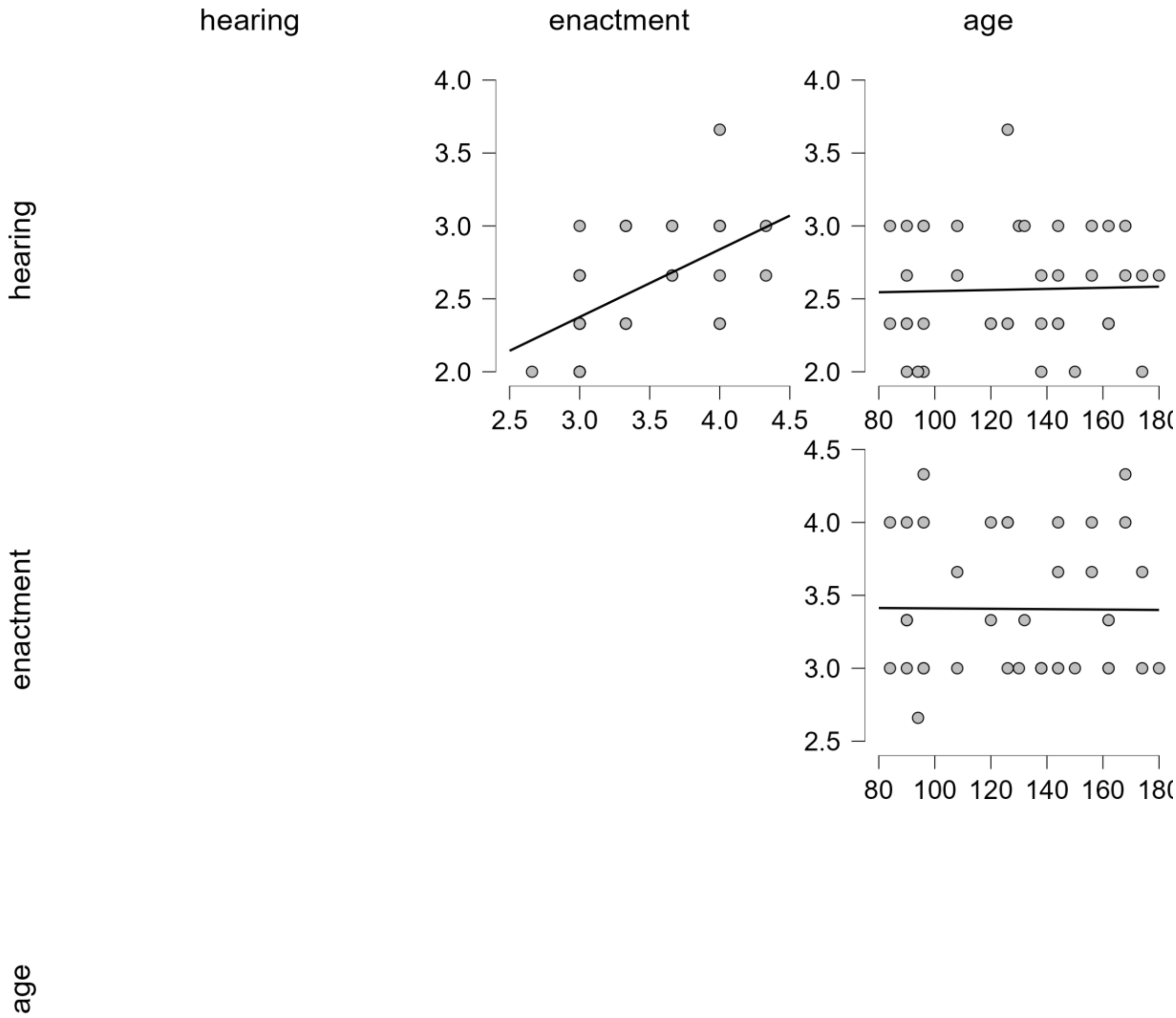
Variable		hearing	enactment	Age
1. hearing	Pearson's r	—		
	p-value	—		
2. enactment	Pearson's r	0.551	—	
	p-value	< .001	—	
3. age	Pearson's r	0.029	-0.009	—
	p-value	0.857	0.958	—

Note. There is a significant relationship between hearing and enactment (r = .551, p < .001). There is no significant relationship for both enactment and age (r=-0.009, p =0.958) and hearing and age (r=.029, p =.857).

Figure 2 shows a possible linear relationship between hearing and enactment. There is no visible linear relationship for both enactment and age and hearing and age.

Figure 2

Scatterplot of the relationships between hearing, enactment and age



Note. The scatterplot illustrates a positive linear relationship between hearing and enactment. The scatterplot illustrates no clear pattern between hearing and age and enactment and age.

Given the fact age is a continuous covariate that might introduce non-linearity, it indicates the assumption of a linear relationship between the dependent and covariate variables might not be met. Therefore, it is crucial to normalize the data before conducting the Repeated Measures ANCOVA. A log-transformation is a possibility to bring outliers and extreme values into a more manageable range (Changyong et al., 2014). By transforming the covariate 'age' and the variables hearing and enactment in JASP, potential distortion could therefore be minimized, meeting the assumption of a linear relationship necessary for conducting the Repeated Measures ANCOVA (Field, 2013).

The relationship between the covariate and the dependent variable has been examined.

Table 9 shows there is a significant relationship for hearing and enactment relationship (*r*=.567, $p <$ .001). There is no significant relationship for both enactment and age (*r*=-0.006, *p* =0.970) and hearing and age (*r*=.046, *p* =.778).

Table 9

Pearsons Correlation for LogHearing, LogEnactment and LogAge

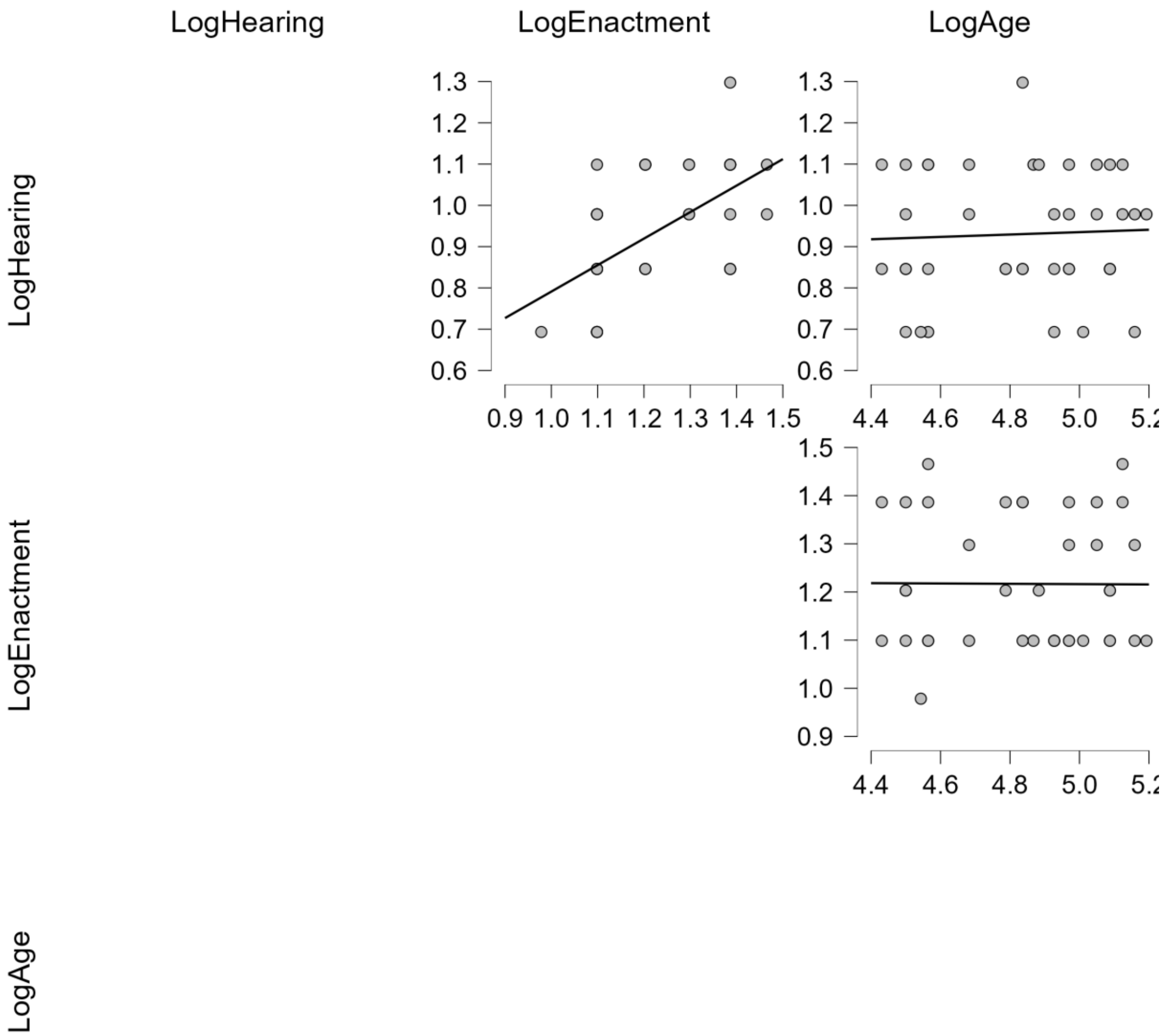
Variable		LogHearing	LogEnactment	LogAge
1. hearing	Pearson's r	—		
	p-value	—		
2. enactment	Pearson's r	0.567	—	
	p-value	< .001	—	
3. age	Pearson's r	0.046	-0.006	—
	p-value	0.778	0.970	—

Note. The Pearson's Correlation coefficient (r) is shown for each pair of variables. There is a significant correlation between hearing and enactment (*r*=.567, $p <$ .001). There is no significant correlation for enactment and age (*r*=-0.006, *p* =0.970) and hearing and age (*r*=.046, *p* =.778).

Figure 3 shows a possible linear relationship between hearing and enactment. There is no visible linear relationship for both enactment and age and hearing and age.

Figure 3.

Scatterplot of the relationships between LogHearing, LogEnactment and LogAge



Note. The scatterplot illustrates a positive linear relationship between LogHearing and LogEnactment. The scatterplot illustrates no clear pattern between LogHearing and LogAge and LogEnactment and LogAge.

Normality

To test for the assumptions of normality of the residuals, a Shapiro-Wilk test was performed. Table 10 shows the descriptive statistics for both the hearing and enactment variable for ASD- and TD children. The Shapiro-Wilk test shows no significant distribution in the hearing condition for TD children (*W*=.905, *p*=.051) and a significant distribution for ASD children (*W*=.869, *p*=.011). There is a significant distribution in the enactment condition for TD (*W*=.888, *p*=.024) and ASD children (*W*=.694, *p*=< .001). There is a significant distribution in age for TD children (*W*=.900, *p*=.041) and no significant distribution for ASD children (*W*=.939, *p*=.234). This indicates there is no normal distribution of the data for both groups in the enactment condition, for ASD children in the hearing condition and for age in TD children. This means the normality assumption is somewhat violated when using these variables.

Table 10

Descriptive Statistics and Shapiro-Wilk Test for Normality of Residuals

	age		hearing		enactment	
	ASD	TD	ASD	TD	ASD	TD
Valid	20	20	20	20	20	20
Missing	0	0	0	0	0	0
Mean	127.300	130.500	2.464	2.664	3.149	3.664
Std. Deviation	28.641	32.216	0.349	0.419	0.333	0.446
Shapiro-Wilk	0.939	0.900	0.869	0.905	0.694	0.888
P-value of Shapiro-Wilk	0.234	0.041	0.011	0.051	< .001	0.024
Minimum	84.000	84.000	2.000	2.000	2.660	3.000
Maximum	180.000	174.000	3.000	3.660	4.000	4.330

Note. There is no significant distribution in the hearing condition for TD children (W=.905, p=.051) and a significant distribution for ASD children (W=.869, p=.011). There is a significant distribution in the enactment condition for TD (W=.888, p=.024) and ASD children (W=.694, p=< .001). There is a significant distribution in age for TD children (W=.900, p=.041) and no significant distribution for ASD children (W=.939, p=.234).

Age, hearing and enactment are variables that might introduce non-normality. Therefore, it was chosen to do a log-transformation to normalize the data by minimizing potential distortion and bringing the values into a more manageable range (Changyong et al., 2014). The log-transformation was performed on hearing (LogHearing), enactment (LogEnactment) and age (LogAge). Table 11 shows the descriptive statistics for the LogHearing, LogEnactment and LogAge variable for ASD and TD children. The Shapiro-Wilks test shows no significant distribution in the hearing condition for TD children (W=.912, p=.068) and a significant distribution for ASD children (W=.875, p=.014).

There is a significant distribution in the enactment condition for TD (*W*=.877, *p*=.015) and ASD children (*W*=.721, *p*=< .001). There is a significant distribution in age for TD children (*W*=.893, *p*=.030), and no significant distribution for ASD children (*W*=.924, *p*=.118). This indicates there is no normal distribution of the data for both groups in the enactment condition, for the ASD children in the hearing and for age in TD children. This means the normality assumption is quite violated as well after doing the log-transformation on the hearing, enactment and age condition.

Table 11

Descriptive Statistics and Shapiro-Wilk Test for Normality of Residuals after log-transformation

	LogAge		LogHearing		LogEnactment	
	ASD	TD	ASD	TD	ASD	TD
Valid	20	20	20	20	20	20
Missing	0	0	0	0	0	0
Mean	4.821	4.841	0.892	0.986	1.142	1.291
Std. Deviation	0.234	0.258	0.141	0.157	0.099	0.125
Shapiro-Wilk	0.924	0.893	0.875	0.912	0.721	0.877
P-value of Shapiro-Wilk	0.118	0.030	0.014	0.068	< .001	0.015
Minimum	4.431	4.431	0.693	0.693	0.978	1.099
Maximum	5.193	5.159	1.099	1.297	1.386	1.466

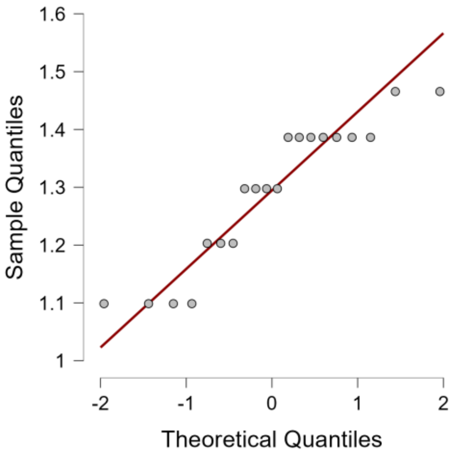
Figure 4 shows it is more likely that the data was more normally distributed after this log-transformation. However, this still might not be enough to ensure the validity of the test. This means the results should be interpreted with caution.

Figure 4

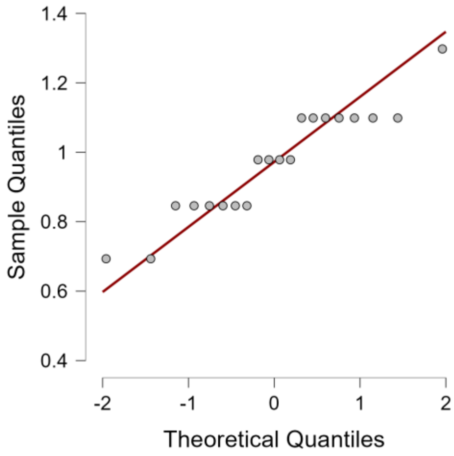
Q-Q plot of LogHearing, LogEnactment and LogAge

4.1 LogHearing

ASD

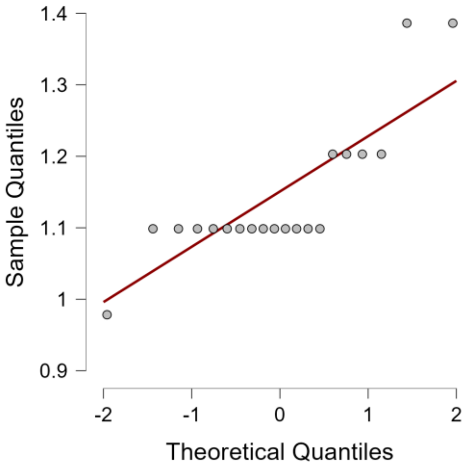


TD

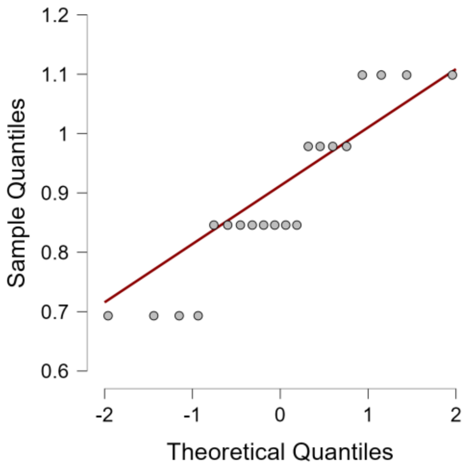


4.2 LogEnactment

ASD

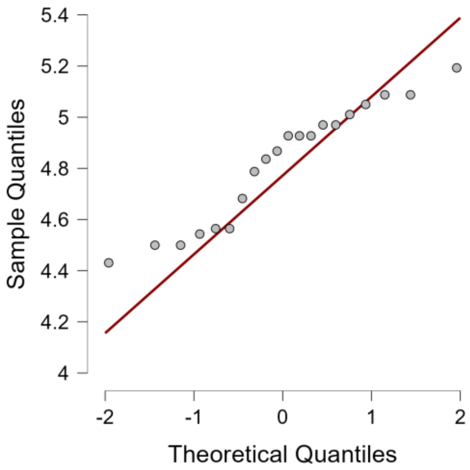


TD

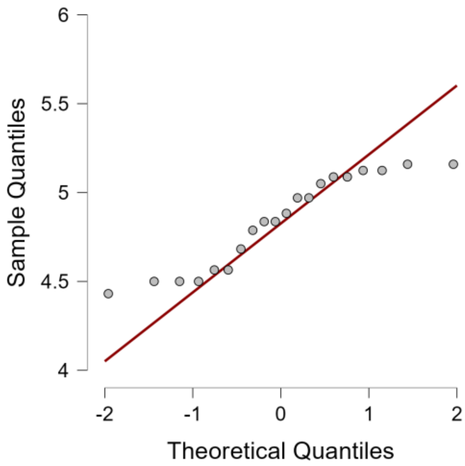


4.3 LogAge

ASD



TD



Equal variances

To test this assumption, a Levene's test of Equality of Variances was conducted. Table 12 shows that there is no difference in the variance (hearing p = 0.681, enactment p= 0.129). Therefore, the assumption of equal variances is not violated.

Table 12

Test for Equality of Variances (Levene's)

	F	df1	df2	p
hearing	0.172	1	38	0.681
enactment	2.406	1	38	0.129

Note. There are no significant differences for the hearing condition (*p*=.681) and the enactment condition (*p*=.129).

Interaction between the covariate and the experimental condition

Table 13 shows instruction ✻ age transformed is not significant (*p*=.661). This means there is no significant interaction between those variables, indicating this assumption was met.

Table 13

Within Subjects Effects of instruction, age and group

Cases	Sum of Squares	df	Mean Square	F	p	η^2_p
Instruction	0.011	1	0.011	1.258	0.269	0.033
Instruction ✻ group	0.027	1	0.027	3.150	0.084	0.078
Instruction ✻ age	0.002	1	0.002	0.195	0.661	0.005
Residuals	0.323	37	0.009			

Note. There is no significant interaction between instruction and age (*p*=.661).

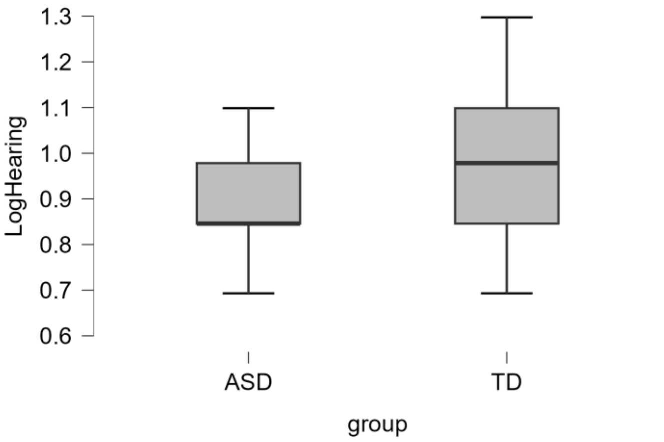
Significant outliers

To test for the assumption of significant outliers, a boxplot was made (figure 5). Figure 5 shows there might be an outlier in the enactment condition for ASD children. In a small sample size, outliers could provide vital information about the population. Removing outliers in small samples may bias the results and can reduce the generalizability and representativeness of the study, which is problematic (Tabachnick, Fidell & Ullman, 2013). To detect the presence of an outlier, the method of the mean plus or minus 3 standard deviations can be used (Howell et al., 1998).

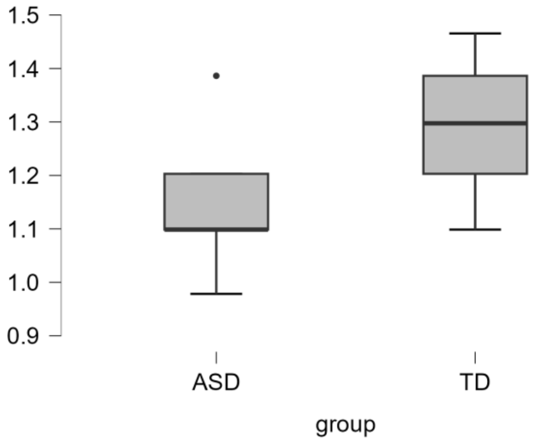
Figure 5

Boxplot of scores for LogHearing, LogEnactment and LogAge in ASD and TD children

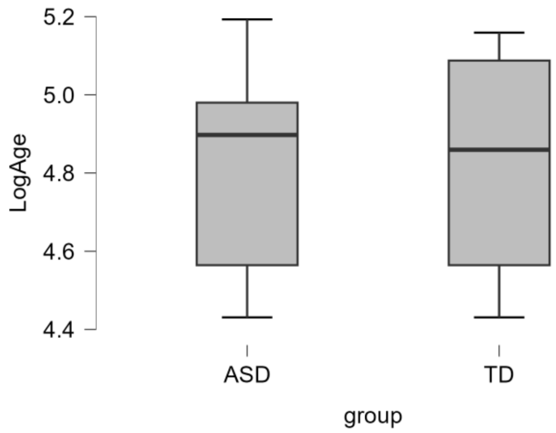
5.1 LogHearing



5.2 LogEnactment



5.3 LogAge



Note. The x-axis represents the group (ASD, TD), and the y-axis represents LogHearing (5.1), Logenactment (5.2) and LogAge (5.3). The boxplots demonstrate the distribution of scores for both groups across the conditions, with lower scores for ASD children in LogHearing and LogAge compared to TD children. An outlier is visible in the LogEnactment condition for the ASD group.

Table 14 shows the descriptive statistics of LogHearing, LogEnactment and Logage for ASD- and TD children. Based on the mean (*M*=1.142) and standard deviation (*SD*=.009) in the enactment condition for ASD children, there is a range with an upper value of 1.315 and a lower value of .469.

The minimum (.987) and maximum (1.386) are within this range, indicating there is no outlier. This means the assumption of significant outliers was met.

Table 14

Descriptive Statistics for Normality of Residuals after log-transformation

	hearing		enactment		age transformed	
	ASD	TD	ASD	TD	ASD	TD
Valid	20	20	20	20	20	20
Missing	0	0	0	0	0	0
Mean	0.892	0.968	1.142	1.291	4.821	4.841
Std. Deviation	0.141	0.157	0.099	0.125	0.234	0.258
Minimum	0.693	0.693	0.978	1.099	4.431	4.431
Maximum	1.099	1.297	1.386	1.466	5.193	5.159

Note. This table illustrates the descriptive statistics for ASD and TD children.