Can instruction through enactment play a role in inclusive education?

Roan Carlier
S2936925
Master Orthopedagogiek
Faculty BSS
Rijksuniversiteit Groningen
Master Thesis
6-12-2024
6669 words

First assessor: Dr. Sarahanne M. Field
Second assessor: Dr. Astrid Menninga

Abstract

In 2013, the Dutch government introduced a law to make mainstream education more inclusive, requiring children with Autism Spectrum Disorder (ASD) and Intellectual Disabilities (ID) to be placed in this educational setting. Both groups face working memory challenges, which hinder their ability to follow verbal instructions. Enactment is a method where verbal instructions are physically enacted, reducing the cognitive load on working memory and potentially making mainstream education more accessible for students with ASD and ID. The current study aimed to determine whether enactment enables students with ASD and ID to perform at the same level as their typically developing peers when receiving only vocal instruction. This research is a re-analysis of data from Xie et al. (2022) and employed a repeated measures ANOVA. The findings indicate a positive effect of enactment on working memory. Both the ASD and ID groups appear to perform as well as, or better than, their typically developing peers when using enactment. However, due to violations of certain assumptions and indications that the ASD and ID groups may not adequately represent their respective populations, the results should be interpreted with caution. Future research should be conducted with larger, more diverse samples.

Samenvatting

In 2013 voerde de Nederlandse overheid een wet in die regulier onderwijs inclusiever maakt, waardoor kinderen met autisme spectrumstoornis (ASS) en intellectuele beperkingen (IB) in deze vorm van onderwijs worden geplaatst. Beide groepen hebben werkgeheugenproblemen, wat hen belemmert in het volgen van verbale instructies. Enactment is een methode waarbij verbale instructie fysiek wordt uitgevoerd. Deze methode is minder belastend voor het werkgeheugen en zou kunnen helpen om regulier onderwijs toegankelijker te maken voor leerlingen met ASS en IB. Het huidig onderzoek had ten doel om te onderzoeken of leerlingen met ASS en IB door enactment op hetzelfde niveau kunnen presteren als hun normaal ontwikkelende leeftijdsgenoten wanneer ze alleen verbale instructie krijgen. Dit onderzoek is een heranalyse van de data van Xie et al. (2022) en gebruikte een repeated measures ANOVA. De resultaten wijzen op een positief effect van enactment op het werkgeheugen. De ASS en IB groepen lijken door enactment even goed of beter te presteren in vergelijking tot hun normaal ontwikkelende leeftijdsgenoten. Door schending van enkele aannames en het feit dat er aanwijzingen zijn voor het feit dat de ASS en IB groep geen afdoende afspiegeling zijn van de populatie, moeten de resultaten voorzichtig worden

geïnterpreteerd. Toekomstig onderzoek zou moeten worden uitgevoerd met grotere, meer diverse steekproeven.

Table of contents

Introduction

In 2013 the Dutch government accepted a law that obligates the Dutch school system to become more inclusive (Ministerie van Justitie en Veiligheid, 2023). This means that, by law, schools are now obliged to offer every student a fitting education. This law results in children with learning disabilities (ID) and autism spectrum disorder (ASD), getting placed in the regular school system. Because schools and their teachers now need to accommodate a much more diverse group of students, teachers must make significant adjustments in their approach to teaching (Tiernan et al., 2018).

Problematic with the inclusion of students with ID and ASD is that they have different educational needs that cannot always be met in regular education. ASD is a neurodevelopmental disorder that is associated with deficits in social interactions, fixed interests, and problems with working memory (WM; American Psychiatric Association, 2013). Because of these problems, students find it harder to adhere to instructions. ID are characterized by significant difficulties in intellectual- and adaptive functioning. The latter is expressed in conceptual, social, and practical adaptive skills (Prelock & Hutchins, 2018; Schalock et al., 2021). Furthermore, ID's are linked with problems in WM (Peng & Fuchs, 2014). Due to these difficulties, both groups of students encounter problems such as lower academic achievements and difficulties in both making and maintaining social relationships (Tiernan et al., 2018; Lüddeckens et al., 2021).

In general education, students are expected to learn by getting verbal instructions that they need to comprehend and put into action, and WM plays a key role in this process (Baddeley et al., 2020). WM describes multiple processes and systems that relate to task-relevant activation, maintenance, and processing of mental information during task performance (Baddeley and Hitch, 1974). The WM can hold a limited amount of information for a limited amount of time (Hartley & Hitch, 2022). Recent research suggests that the average adult can hold about three to five items when they are not rehearsed (Cowan, 2010). During life span, the WM develops so that students can perform increasingly complex tasks (Baker-Ward et al., 2021). Between the ages of 4 and 11, the WM capacity increases gradually (Alloway et al., 2006), and specific parts of the WM can even develop up until puberty (Luciana et al., 2005). When a student hears a spoken instruction the WM holds the information so the student can execute the complex cognitive action that enables them to adhere to the instruction (Jaroslawska et al., 2016).

The problems that ASD and ID students encounter with regard to WM are similar (Peijnenborgh et al., 2016; Wang et al., 2022). Research suggests that the problems most often are centered in either the central executive (Ring et al., 2020; Henry, 2010) or the episodic buffer (Maras et al., 2012; Hronis et al., 2017).[1] These problems in general result in an overburdened WM (Patrick, 2020). It is important to note that even a mild impairment in WM can have a significant effect on the academic achievements of a student. Research even suggests that WM capacity is a reliable predictor for academic achievement (Alloway & Alloway, 2010). Because of these problems with WM, students with ASD and ID find it harder to adhere to spoken instruction than their typically developing (TD) peers. Therefore, students with ASD and ID show a different approach to learning (Kim & Kasari, 2023). When students are placed in special education they benefit from the neutral learning environment and when tasks are given in chunks (Martin, 2009; Ben-Arieh & Miller, 2009). For successful inclusive education, it is important that teachers have access to a method that accommodates both TD students and students with additional educational needs (De Boer et al., 2010).

An effective way to meet the educational needs of all students in regular education could be by making use of enactment. Enactment is the process in which an instruction is not only listened to but also physically acted out (Xie et al., 2022). Research shows that, as long as the tasks are not overly complex, enacted action phrases are better retained (Xie et al., 2021). When enacting the action phrases visuomotor processes are activated and the WM functions better (Allen et al., 2019), this improvement in WM functioning is called the enactment effect (Waterman et al., 2017).

Enactment reduces cognitive strain on WM by integrating movement information (Xie et al., 2021). Physical movement is fundamental in learning throughout life (Mavilidi et al., 2019). In this process, the body functions as a constituent of the mind (Macrine & Fugate, 2022). Consequently, the use of movement and gestures can significantly enhance learning while reducing the cognitive load (Agostinho et al., 2015; Mavilidi et al., 2015). Gestures, in particular, support the WM in two ways. Firstly, physical gestures act as a kind of placeholder for short-term memory; they can function to offload mental information (Cook, 2013). The gestures will make sure that the decay of memory gets postponed by which the

[1] The WM consists of four parts with their own function. Firstly, the information enters the memory in the central executive which orders information. Secondly, visual information goes to the visuospatial sketchpad and verbal information goes into the phonological loop. Finally, the episodic buffer integrates information form the visuospatial sketchpad and the phonological loop into a coherent whole. (Baddeley and Hitch, 1974)

demand on the WM gets reduced. Secondly, the used gestures can serve as an embodiment of the problem, therefore attention can be focused elsewhere (Sepp et al.,2019). By combining these physical gestures with existing knowledge, enactment maximizes the use of all components of the WM without adding cognitive load (Yang et al., 2013).

Because enactment puts less of a strain on the WM it is possible that it may help students with ASD and ID to perform in a similar way as TD students, resulting in fewer problems for the students with ASD and ID. Hypothetically, teachers could then use one universal form of instruction for the entirety of the class. This might generally improve students' adherence to instructions given by the teacher, not just for children with developmental challenges. Such benefit goes beyond improving the prospect of inclusive teaching and can contribute to more cooperative classrooms, especially for groups of children below the age of 11 for whom WM capacity is relatively limited.

Research questions

The current study aims to determine if enactment can help ensure students with ASD and ID can participate in regular education. For this study, it will be assumed that the WM performance of TD students when given verbal instruction is sufficient to successfully participate in regular education. Therefore the research question guiding this paper is:

" Can the use of an enactment instruction method improve the WM performance of students with ASD and ID, such that they perform similarly to their TD peers when given verbal instruction?"

Accordingly, the following is hypothesized:

H_0: The average span score for WM performance of children with ID or ASD when receiving instruction through enactment is significantly lower than the average span score for WM performance of TD children when receiving only verbal instruction.

H_A: The average span score for WM performance of children with ID or ASD when receiving instruction through enactment is equal or significantly higher than the average span score for WM performance of TD children when receiving only verbal instruction.

Method

For this quantitative research, the data gathered by Xie et al. (2022) will be reanalyzed. The original study used a mixed design and explored if the effect of enactment on WM for instructions differs between ASD-, ID- and TD students. All students, aged 7 to 15

years old, were asked to orally recall instruction sequences in three conditions: vocal instruction, imagining enactment, and enactment.

Design

The current study has an experimental between-subjects design with three groups and two conditions. The two conditions are vocal instruction (hearing) and instruction through enactment (enactment). The dependent variable is the score on the WM task, the span score. A span consists of at least three correctly recalled sequences, for each span one point was assigned. The items per sequence increased when the span was correctly completed and started with one item per sequence. The span score was determined by the following formula: Span score = $N + n * 0.33$. In the formula, N = the number of passed spans, and n = the number of correct sequences in the failed span. The independent variable is the developmental category (TD, ASD, and ID) and whether or not the students were asked to use enactment or only rely on hearing verbal instruction.

Participants

The minimum number of participants for the study is 51, with each group needing a minimum of N=17. This was determined by a power analysis performed by Xie et al. (2022). In total, the sample size of the current study is N = 60, divided into three groups (i.e. 20 students per group).

The first group consists of 20 participants with an ID, four female and 16 male. These participants did not meet the criteria for ASD. If a student, in addition to ID, also met the criteria for attention-deficit/hyperactivity disorder (ADHD), pervasive developmental disorder not otherwise specified, Down syndrome, or other etiologies they were excluded from the experiment. The second group consists of 20 participants with ASD, four female and 16 male. All participants in this group were clinically diagnosed with ASD by a specialized doctor. Exclusion criteria were: a history of epilepsy, neurological abnormalities, significant head injury, psychosis, anxiety disorder, or ADHD. The third and last group consists of 20 TD students, four female and 16 male. The TD students have no history of any neurological or psychiatric diagnoses. All students that participated in the study were right-handed. The age of all participants varies between 7 and 15 years old. Their age was measured in months due to the effect of age on WM. There were no group differences in age.

All the students in the ASD and ID groups were recruited from special education schools that participated in the original experiment. The parents of these students were sent a written informed consent document after which they gave consent for their child's

participation in the experiment. The original paper does not state how the TD participants were recruited and how and if parents were informed.

Procedure

Pre-testing

Before the start of the main experiment, the crystalized and fluid intelligence of all students was tested. The Raven Standard Progressive Matrices Test (SPM; Raven et al, 1998) was used to measure fluid intelligence and the Peabody Picture Vocabulary Test (PPVT-R; Sang & Miao, 1990) was used to measure crystallized intelligence. Students could score either low, medium, or high on both intelligence scales. For inclusion, a student's scores on both intelligence tests needed to be equal. By ensuring that the intelligence profiles are balanced, the authors can more confidently attribute the WM deficits to either ASD or ID. All TD and ASD students scored either high or medium and the ID students all scored low on both tests.

Finally, the students were tested on the severity of the ASD symptoms. The two instruments that were used are the Childhood Autism Rating Scale (CARS-CV; Schopler et al., 1980) and the Autism Behavior Checklist (ABC-CV; Yang et al.,1993). All students in the ASD group met the diagnostic criteria for ASD and their scores were higher than those of the students in the other groups. None of the other students met the diagnostic criteria for ASD.

Materials

During the original experiment, the following 14 objects were used to give instructions: scissors, key, pencil, comb, towel, cup, tissue, umbrella, spoon, glove, chopsticks, mirror, mask, and toothbrush. These 14 objects were paired with the following six actions: spin, touch, push, shake, knock, and turn over. All objects and movements were paired in various combinations to form action phrases. Items could be repeated in one sequence but actions could not.

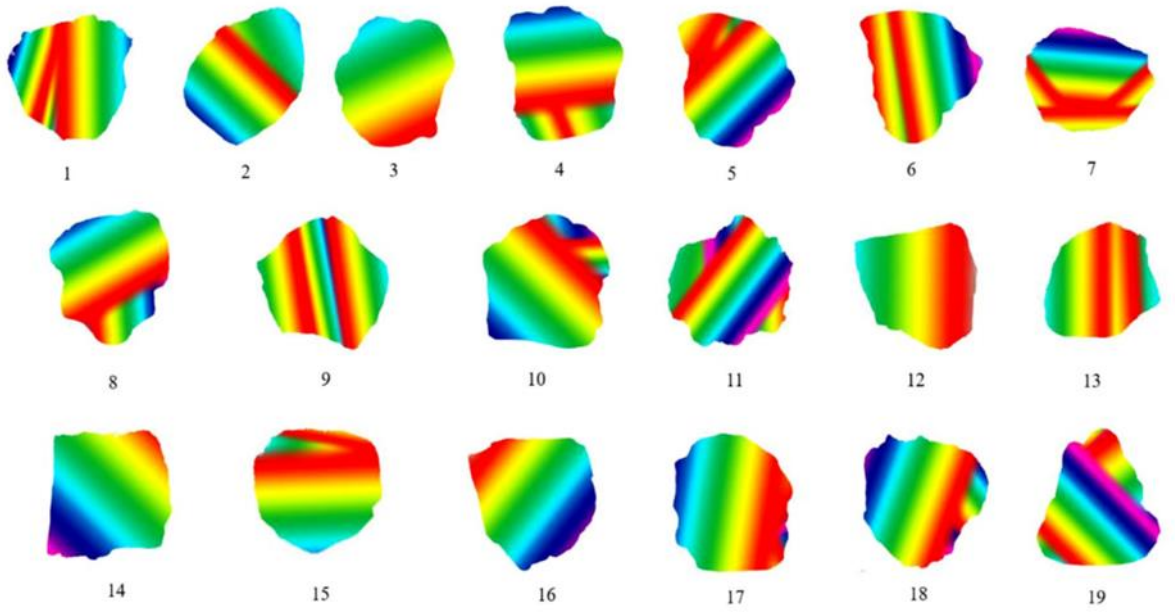
Study procedures

Before the start of the experiment, the participants were given time to familiarize themselves with the different objects used in the experiment. The participants were given prerecorded action phrases to get used to the instructions. In the real experiment, the action phrases were given by an on-site and trained experimenter. In the experiment, the participants were given a period of three seconds to act out the instructions. In the vocal instruction

condition, the participants were shown a colorful irregular figure (Figure 1) during these three seconds. These figures differed after each instruction in a sequence.

Figure 1.

Colorful irregular figures used in the hearing condition



Note: From ''*Can Enactment and Motor Imagery Improve Working Memory for Instructions in Children whith Autism Spectrum Disorder and Children with Intellectual Disability?*'', by Xie et al., 2022 *Journal of Autism and Developmental Disorders*, *54*(1), 131–142. (https://doi.org/10.1007/s10803-022-05780-z)

In the original experiment, all participants completed three conditions, namely a hearing condition, a motor imagery condition, and an enactment condition. The order in which these three conditions were completed varied. As the current study only focuses on the enactment and the hearing condition, only these will be further described.

The main experiment was conducted as follows. Participants were presented with a sequence of spoken instructions. These were presented sequentially with a three-second interval. This interval was silently counted out by the experimenter. In the hearing condition, the participants were asked to watch a colorful irregular figure during these three seconds. In the enactment condition, the participants were asked to act out the instruction with their hands. After the entire sequence, the participants were asked to orally recall the sequence of instructions in correct order, within one minute. If, after four sequences, the participants correctly recalled at least three sequences, the next sequence was presented. If the participant did not meet this criterion, the test was stopped.

During the enactment condition, all objects in the current sequence were presented in an arbitrary order until the entire sequence was completed. After completing the sequence new objects were introduced. Participants were only allowed to touch the objects when they were in the enactment condition. In each condition, the tasks were demonstrated to the participants.

Reliability and validity

The reliability and validity of the tests used in the original study can be found listed below.

Raven Standard Progressive Matrices Test

The SPM test has been evaluated by the COTAN (the Commissie Testaangelegenheden Nederland) and has been rated sufficient (Egberink & Leng, 2014). According to the COTAN manual, this rating indicates that the instrument is suitable for professional use (Evers et al., 2010). The materials of the test are in good shape and the procedure is clear. The reliability of the SPM test is sufficient, which means that Cronbach's alfa (α) and Guttman's lambda2 (λ) are between 0.80 and 0.90. However, the data indicates it is less dependable for the extreme scores (Egberink & Leng, 2014). The construct validity of the test was found to be sufficient (Egberink & Leng, 2014). This indicates that the constructs are theoretically grounded and that the correlation coefficient between the test and related constructs is between r=0.30 and r=0.50. (Evers et al., 2010). The criterion validity was found to be insufficient because not enough research was conducted on this topic (Egberink & Leng, 2014). It is important to note that this evaluation is of the Dutch version of the SPM test, results could differ for the Chinese version. The COTAN manual states that any translation needs new research (Evers et al., 2010). To the best of the researchers' knowledge, no recent data was available for the Chinese version of the SPM test.

Peabody Picture Vocabulary Test

The PPVT-R test has been evaluated by the COTAN and has been rated sufficient (Egberink & Leng, 2005), which means that the test has been found suitable for professional use (Evers et al., 2010). The materials and instructions are good and clear, but the norms show to be outdated (Egberink & Leng, 2005). The reliability of the test is good, which means that Cronbach's alfa (α) and Guttman's lambda2 (λ) are above 0.90 (Egb erink & Leng, 2005). The construct validity is sufficient (Egberink & Leng, 2005). This indicates that the constructs are theoretically grounded and that the correlation coefficient between the test and related constructs is between r=0.30 and r=0.50. The criterion validity was found to be

insufficient because no research was conducted on this topic (Egberink & Leng, 2005). It is important to note that this evaluation is of the Dutch version of the PPVT-R test, although the Dutch test evaluation is insightful, results could differ for the Chinese version. To the best of the researchers' knowledge, no recent data was available for the Chinese version of the PPVT-R.

Childhood Autism Rating Scale

According to research by Schopler et al. (1980), the CARS test is a reliable instrument to test for the symptoms of ASD. The internal consistency was tested with a Cronbach's alfa and scored α=0.94, which is a high score. The inter-rater reliability has a correlation of r=0.71. The test is based on direct behavioral observation, which makes it very useful for in-depth behavior descriptions (Schopler et al., 1980). The test can successfully distinguish two groups of children with different profiles of ASD (Schopler et al., 1980). According to other research, the CARS is reliable and valid for all age groups (Garfin & McCallon, 1988). The CARS seems to be less successful in distinguishing children with different backgrounds (Schopler et al., 1980). It is important to note that this evaluation is of the English version of the CARS, although insightful, results can differ for the Chinese version. To the best of the researchers' knowledge, no recent data was available for the Chinese version of the CARS.

Autism Behavior Checklist

When testing for the reliability of the ABC mixed results are found. The total test is found to be reliable (α= 0.89). However, when all sub-scales are tested separately Cronbach's alpha is below α=0.8, which is not sufficient for screening purposes (Eaves & Williams, 2006). This indicates that the total score can be insightful but the sub-scales are not sufficiently reliable. The concurrent validity of the ABC differs from good (r=>0.80) to insufficient (r=<0.30) (Eaves & Williams, 2006). Because of the varying findings, results should be handled with caution. Cassidy (2013) states that when the ABC is used in conjunction with other diagnostic instruments and methods, it can be useful as a symptom inventory, to be used by clinicians in structuring their evaluation. It is important to note that this evaluation is of the English version of the ABC, although insightful, results can differ for the Chinese version. To the best of the researchers' knowledge, no recent data was available for the Chinese version of ABC.

Data analysis

The data that resulted from the earlier described experiment was reanalyzed in JASP (version 0.19.1) (JASP Team, 2024), a program for statistical analysis.

Statistical testing

To answer the research question the data was analyzed using a 3x2 contrast coded repeated measures ANOVA. This test compares the means of groups and tests significance (Field, 2013). In this case, the dependent variable was the span score for WM. The independent variables are the participant group (TD, ID, and ASD) and condition (vocal instruction and enactment). When testing the assumptions (Appendix A) for a contrast coded repeated measures ANOVA, the assumptions about the normality of the residuals and significant outliers were violated. Therefore, a log transformation was performed on the span scores for both the hearing- and enactment conditions. A log transformation can help by correcting for influential outliers and making the distribution of the variables more normal (Feng et al., 2014). However, when controlling the assumptions again with the log transformed variables, there were more influential outliers, and the Shapiro-Wilk test for normality remained insignificant (Appendix A). According to research conducted by Feng et al. (2014), in such cases, it is preferable to use the original data and approach the findings with caution.

To test for the influence of age, an ANCOVA analysis was planned. However, when the data was tested for the assumptions for ANCOVA (Appendix B) it became apparent that there was no linearity between the covariate and the dependent variable. Because this linearity forms the base of the validity of the test (Kirk, 2013), it was not possible to conduct an ANCOVA. Therefore, Spearman's correlation, which has less strict requirements for linearity but can still indicate a relationship between age and span score, was conducted. The assumptions of the Spearman's correlation are described in Appendix C.

All statistical analyses were conducted with a significance level set at $p = 0.05$, indicating that results with a p-value below this threshold were considered statistically significant. To determine effect size, η^2_p was used for the main effect of the repeated measures ANOVA, and Cohen's d was used for the contrasts. The η^2_p is a measure for effect size that corrects for effects of other variables, it represents the proportion of the variation that can be explained by a specific factor (Norouzian & Plonsky, 2017). A $\eta^2_p = 0.01$ is believed to be a small effect, a $\eta^2_p = 0.06$ is believed to be a medium effect and a $\eta^2_p = 0.14$ and above is believed to be a strong effect (Richardson, 2011). Cohen's *d* is a measure of effect size. A Cohen's $d = 0.2$ is a small effect, a Cohen's $d = 0.5$ is a medium effect, and a Cohen's $d = 0.8$ and above is a large effect (Cohen, 2013).

Missing data

In the original study by Xie et al. (2022) it is not described if there was any missing data, and if there was any missing data, why it was missing, and how it was handled. Upon reflection, the best method for handling missing data would be multiple imputation. This method would be the most fitting because it can be used on smaller data sets and preserves data (Van Buuren, 2018).

Results

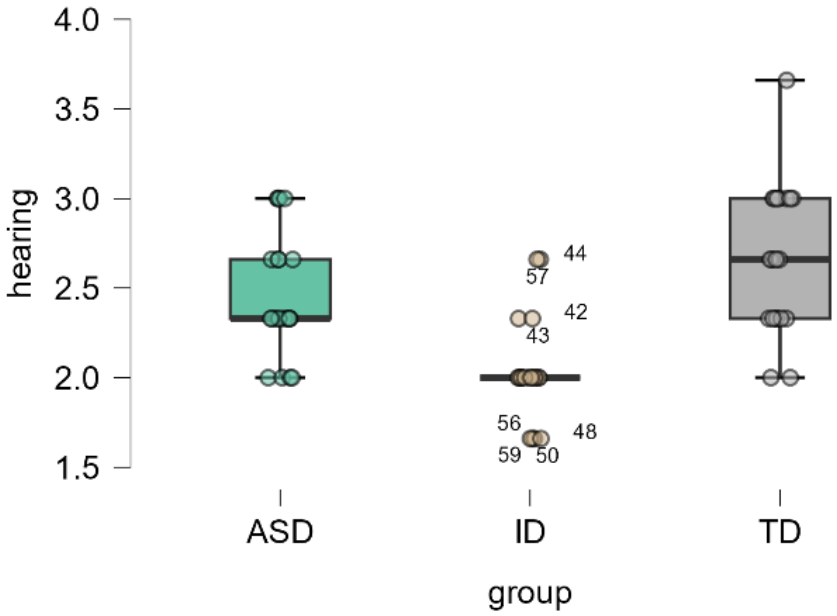
The research question of the current research is: Can the use of an enactment instruction method improve the WM performance of students with ASD and ID, such that they perform similarly to their TD peers when given verbal instruction? All participants were able to complete both the hearing condition and the enactment condition. To answer this question a 3x2 contrast coded repeated measures ANOVA was conducted, the assumptions check is described in Appendix A. The main effect of instruction type was found to be significant $F(1, 42)= 236{,}471$ $p=<0.001$ $\eta^2_{p=}$ 0.467. It is important to note that, due to violated assumptions, all results should be approached with caution.

Hearing condition

The spread of all span scores is plotted in boxplots (Figure 2). In the boxplot of the ASD students (Figure 2) it is shown that there are multiple upper-bound adjacent scores, but

Figure 2.

Boxplots of the spread of the span scores in the hearing condition



Note. Analysis performed on data from "*Can Enactment and Motor Imagery Improve Working Memory for Instructions in Children with Autism Spectrum Disorder and Children with Intellectual Disability?"* by Xie et al. (2022) *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-022-05780-z

no significant outliers are labeled. In the boxplot of the ID students (Figure 2) it is shown that the spread is narrow and that there are a lot of outliers labeled. The boxplot of the TD students (Figure 2) shows three adjacent scores and no outliers.

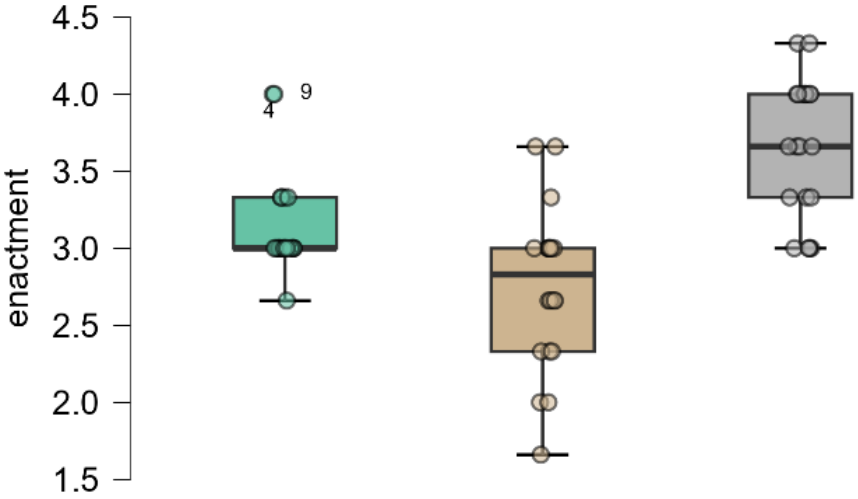
In the hearing condition the participants with ASD scored as follows; *M*= 2,464 (*SD*= 0,349). The participants with ID had an average score of *M*=2,031 (*SD*= 0.284), and the participants with TD had an average score of *M*= 2.664 (*SD*= 0.419). The difference between the group of participants with ASD and the group of TD participants was not significant (*t*= -1.576, *p*= 0.118, Cohen's *d*= -0.498). The group of participants with ID scored significantly lower than the group of TD participants (*t*= -4.998, *p* <.001, Cohen's *d*= -1.557).

Enactment condition

In Figure 3 the spread of all groups in the enactment condition is displayed in boxplots. In the boxplot of the ASD students it is shown that there are multiple adjacent scores and labeled outliers. In the boxplot of the ID students it is shown that the scores are spread out more compared to the other groups. The boxplot of the TD students shows a tighter spread without outliers but with multiple extreme scores.

Figure 3.

Boxplots of the spread of the span scores in the enactment condition



Note. Analysis performed on data from "*Can Enactment and Motor Imagery Improve Working Memory for Instructions in Children with Autism Spectrum Disorder and Children with Intellectual Disability?"* by Xie et al. (2022) *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-022-05780-z

The ASD students had an average span score of *M*= 3.149 (*SD*= 0.333), and the scores in the hearing condition are were significantly lower than their score in the enactment condition (*t*= -7.597, *p*=<.001, Cohen's *d*= 1.707). The group of participants with ID had an average score of *M*=2.747 (*SD*= 0,528) in the enactment condition. The scores of this group

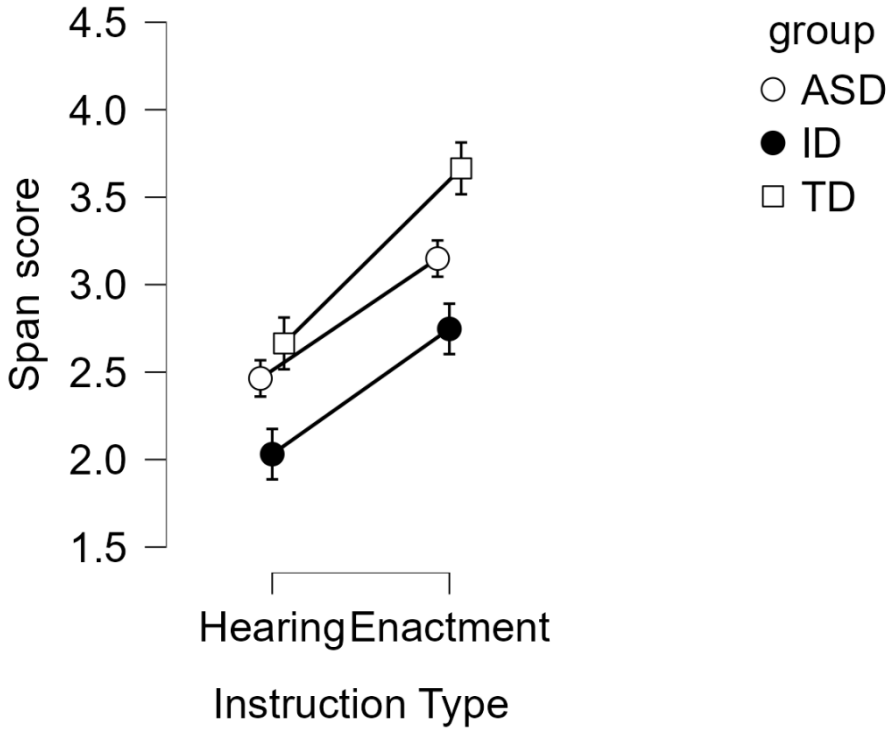
in the hearing condition are significantly (*t*= -7.941, *p* <.001, Cohen's *d*= 1.784) lower than in the enactment condition. The group of TD participants scored an average of *M*=3.664 (*SD*= 0,446). The score in the hearing condition was significantly lower (*t*=-11.096, *p* <.001, Cohen's *d*= 2.493). In descending order, the effect size of the groups was, TD (Cohen's *d*= 2.493), ID (Cohen's *d*=1.784), and last ASD (Cohen's *d*=1.707).

Main question

The first contrast of interest was between TD students in the hearing condition and both ASD and ID students in the enactment condition. In Figure 4 the mean span scores across all three groups are illustrated. The students in the enactment condition scored significantly higher than the TD students in the hearing condition (*t*= 2. 684, *p*= 0.010, Cohen's *d*= 1.415).

Figure 4.

Descriptive plot – differences in all groups between the hearing and the enactment condition



Note. Analysis performed on data from "*Can Enactment and Motor Imagery Improve Working Memory for Instructions in Children with Autism Spectrum Disorder and Children with Intellectual Disability?"* by Xie et al. (2022) *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-022-05780-z

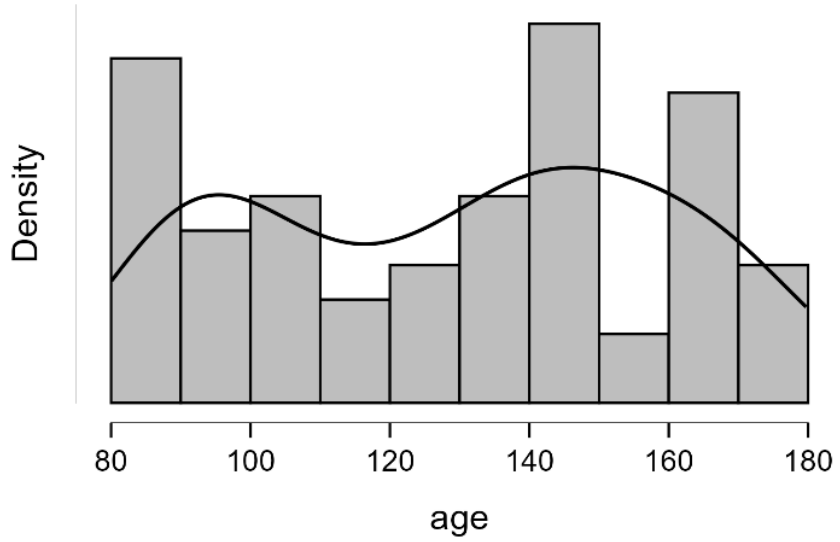
The second contrast tested is between TD students in the hearing condition and ASD students in the enactment condition. The ASD students in the enactment condition scored significantly higher than the TD students in the enactment condition (*t*= 3.822, *p*= 0.001, Cohen's *d*= 1.209).

The third contrast that was tested was between TD students in the hearing condition and ID students in the enactments condition. As mentioned, the mean span score of the ID group in the enactment condition was *M*=2.747, this score is higher than the score of the TD group in the hearing condition which was *M*=2,664. This difference, however, was not found to be significant (*t*=0.516, *p*=0.516, Cohen's *d*= 0.207).

Influence of age

Figure 5.

Spread of age in months over all groups



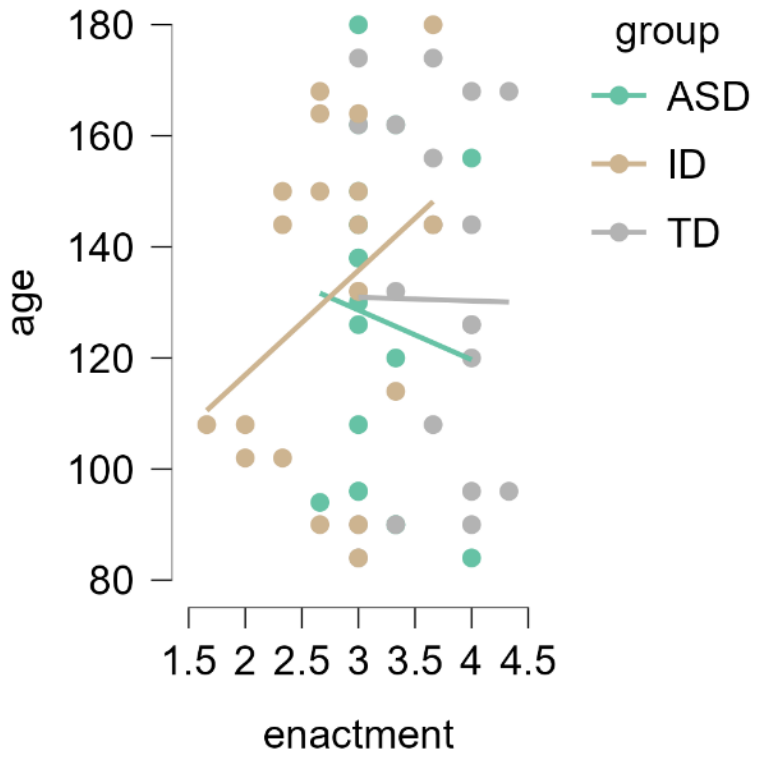
Note. Analysis performed on data from "*Can Enactment and Motor Imagery Improve Working Memory for Instructions in Children with Autism Spectrum Disorder and Children with Intellectual Disability?"* by Xie et al. (2022) *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-022-05780-z

In Figure 5 the spread of age in months over all groups is plotted. It is shown that age is not normally distributed in the sample and there are two gaps visible in the spread.

To test for the influence of age, a Spearman correlation was conducted, the assumptions are described in Appendix C. Firstly, all scores of all students in the enactment were tested for a correlation with age. The scatterplot in Figure 6 illustrates the relationship between enactment and age for all groups. The correlation between age and span scores of all three groups in the enactment condition was Pearson's ρ=0.030, and this was not found to be significant (*p*=0.409). When looking at the spread of the scatterplot, there is no clear positive relation visible. When analyzing all groups separately, the following was found. In the ID group, the correlation was Pearson's ρ= 0.235; this indicates a small positive relationship but it was not found to be significant (*p*=0.159). In the ASD group, the correlation was Pearson's ρ=-0.139; this indicates a small negative relationship, but it was not found to be significant

Figure 6.

Scatterplot – Age and span score in the enactment condition



Note. Analysis performed on data from "*Can Enactment and Motor Imagery Improve Working Memory for Instructions in Children with Autism Spectrum Disorder and Children with Intellectual Disability?"* by Xie et al. (2022) *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-022-05780-z

(p=0.721). Lastly, in the TD group the correlation was Pearson's ρ =0.006; this indicates a small positive relationship, but it was not found to be significant (p=0.490).

Conclusion and discussion

This research attempted to identify if enactment-based instruction can enhance WM performance in students with ASD and ID to match the performance of TD students who received the regularly used verbal instruction. The main effect of instruction type was found to be significant. This means that all scores in the current study were significantly higher with instruction through enactment than with verbal instruction. The effect size that was found is considered to be a large effect size. These findings are in line with the expectation that WM performance increases with enactment because it puts less of a strain on the WM. When comparing the span scores of the ASD- and ID students in the enactment condition to the span scores of the TD students in the hearing condition, the current study found indications that enactment does in fact help students with ASD and ID match the performance of TD students.

Hearing condition

In the hearing condition it was expected that both the ID- and the ASD students would score significantly lower on the span-test than the TD students because research shows that both groups of students have a limited WM capacity (Colley, 2006); Peijnenborgh et al., 2016). The group of ID student did, in line with the expectations, scored significantly lower than the TD students in the hearing condition. Based on these findings, it can be concluded that the current study implicates that the ID students do in fact have a significantly smaller WM performance.

However, the group of ASD students did not score significantly lower than the TD students. As mentioned above this is not in line with the expectations based on the literature. It is possible that the findings deviate from the existing literature due to the small sample size. The limited sample size increases the influence of outliers, and the spread of the data is suboptimal, which may have affected the results.

A second possible explanation is that, all included ASD students had a medium or high intelligence. This may not be a good representation of the population of ASD students, because research shows that ASD occurs across all levels of intelligence (Hollander et al., 2018). Furthermore, it is also possible that the ASD students in the sample have developed strategies that help them strengthen their WM that their less intelligent peers do not possess. If the ASD students in the sample use such strategies it is possible that the span-test does not measure the WM performance of the average ASD student.

In addition, previous research shows that ASD students, their parents and their teachers report that they do in fact experience problems with administering to spoken instruction (Jaroslawska et al., 2016; Shurr et al., 2021). Based on the above-mentioned limitations and the findings in multiple previous studies we will presume that the students with ASD do in fact experience significant difficulties with following spoken instruction.

Enactment

In the enactment condition, all students of all groups performed significantly better than in the hearing condition. This indicates that all students had a larger WM capacity when using instruction through enactment. This is in alignment with the findings of previous studies (Wang et al., 2022; Xie et al., 2021).

All found Cohen's *d* scores are considered to reflect very strong effects with clear practical implications. This aligns with the findings from previous studies that found that WM performance increases with the use of enactment (Allen et al., 2019; Xie et al., 2022).

Enactment reduces the load on WM by integrating movement and gestures, which act as placeholders to offload the burden upon the WM (Sepp et al., 2019).

Main question

To answer the current research's main question the results of the ASD students and the ID students in the enactment condition were compared to the results of the TD students in the hearing condition. When comparing both the ASD and the ID students to the TD students, results showed that the students in the enactment condition scored significantly higher than the TD students in the hearing condition. The effect size is considered to be very strong which means that the findings have a strong practical implication. This implies that, instructions through enactment, in fact, can help to support these students improve their WM capacity.

When separating the two groups of students it becomes apparent that the effect size in the ASD group and the ID group strongly differ. The ASD students, as hypothesized, had a significantly higher score in the enactment condition than the TD students in the hearing condition. The values found strongly suggest that the implementation of enactment in regular education can indeed support students with ASD to improve their WM performance.

The group of ID students had an average span score that was similar to the average span score of the group of TD students. However, statistical analysis showed that the group of ID students did not score significantly better than the TD students. This means that based on the current study, we cannot conclude that enactment supports ID students to such a degree that they perform similarly to TD students in regular education. This is in contradiction with the hypothesis and can be declared in several ways.

Firstly, it is possible that the findings are not significant because the spread of the scores of the ID students in the enactment conditions is relatively large compared to the other groups. This can have a negative effect on the significance. It is important to note that the population of ID students is highly heterogeneous in practice, which may account for the larger spread.

Second, it is possible that the WM problems the ID students have, differ from the problems the ASD students have. Thurm et al. (2019) found that WM problems in ID students are often more diffuse, whereas in ASD students, these difficulties may manifest in specific contexts, such as in social situations or when planning and organizing activities. Therefore, enactment might have a different effect on ID students.

Third, it is important to note that all participants in the ID group exhibit low intelligence, whereas participants in the ASD and TD groups display either medium or high intelligence levels. According to Thurm et al. (2019), the WM problems in ID students are often linked to their intelligence level. When taking this into account, it is possible that the differences in WM performance between the groups were too big to begin with. Specifically, students with ID may have had a WM performance that was considerably lower than the WM performance of the ASD and TD students. This means that even if enactment would cause a significant improvement in the WM performance of the ID group, it could still not be enough to make them perform to a similar or higher level as the TD group without using enactment.

Lastly, research shows that the problems ID students have with verbal instruction might not only link to an impaired WM but can also be linked to significant deficits in language (Hicks et al., 2015) and planning (Klefbeck, 2021). Based on the findings in the current study and the above-mentioned findings, it may be possible that the use of enactment is insufficient for the ID students.

Influence of Age

Based on the literature, that states that WM performance increases naturally during life (Baker-Ward et al., 2021), with a peak at around 11 years old (Alloway et al., 2006), age was expected to correlate positively with enactment. In contradiction to the expectations, age did not significantly correlate with the span score in the enactment condition. In the literature, there is a lot of proof for the effect of age on WM. Therefore is it possible that the characteristics of the sample made it harder to pick up on the effect of age.

It is possible that the sample could not capture the effect of age on the WM because of the small sample and the fact that the spread of age throughout the sample is suboptimal. There is little data and there are relatively few participants at the age of (132 months) and relatively many participants at around 80, 140, and 160 months. Because research shows that there is a fast improvement of WM in childhood with a peak in childhood that is equal between girls and boys (Ahmed et al., 2022), a smaller, well-rounded sample should be able to capture the effects. However, Figure 5 showed that the age in the current sample is not evenly distributed which could explain why the effect of age could not be measured.

Based on previous research it can be concluded that, even though the current research was not able to test for the effect of age, age does in fact play a role in the WM development. Therefore, any conclusions that are drawn based on the current research must be handled with caution.

Limitations

The findings of this study are difficult to generalize due to several limitations of the study. The first limitation is the relatively small sample size, which increases the impact of variations in the data. This makes it difficult to find significant results and draw meaningful conclusions from them.

Secondly, the sample includes relatively few girls (four girls in every group). Since previous research has shown significant differences in how autism manifests in males versus females (Rivet & Matson, 2011), it cannot be stated with certainty that the findings can be generalized to the entire population.

A third limitation of the study is that the sample of ASD students may not reflect the population. As previously discussed the ASD students in the sample all have either medium or high intelligence, while the ASD students in the population also have low intelligence.

Another limitation of the study is that it was conducted in China with Chinese students. Due to significant cultural differences between other countries and China, as well as the limited description of the schools from which the students were included, it cannot be stated with certainty that the findings can be generalized to other contexts.

Fourth, there is no recent data on the reliability and validity of the Chinese version of the questionnaires used to select participants. Therefore, it is possible that these older versions of the used questionnaires no longer align with contemporary expectations of children and their abilities. As a result, the reliability and validity of these Chinese instruments cannot be guaranteed.

Finally, the original study by Xie et al. (2022) did not report on two important parts of how the research was conducted. They do not describe how they recruited the TD students and they did not report on getting consent from their parents. In addition, the original research does not describe whether they had missing data, and if so, how they handled it. Therefore, conclusions drawn from the data and their implications should be approached with great caution.

Recommendations

When taking the limitation of the current study into account, we can still conclude that the findings show clear signs of a positive effect of instruction through enactment on all students when compared to only verbal instruction. Therefore, it could be wise to replicate the study with students from other countries and cultures, a larger sample, more girls, and a better distribution of age. This can help with the generalizability of the findings.

As the findings suggest that enactment has a large effect on all groups of students it is also recommended that future research focusses on the practical implementation of instruction through enactment in regular education.

Finally, future research could further look into the additional needs of ID students, so enactment has the maximal effect on every group of students in regular education.

References

Agostinho, S., Tindall-Ford, S., Ginns, P., Howard, S. J., Leahy, W., & Paas, F. (2015). Giving learning a helping hand: finger tracing of temperature graphs on an iPad. *Educational Psychology Review*, *27*(3), 427–443. https://doi.org/10.1007/s10648-015-9315-5

Ahmed, S. F., Ellis, A., Ward, K. P., Chaku, N., & Davis-Kean, P. E. (2022). Working memory development from early childhood to adolescence using two nationally representative samples. Developmental Psychology, 58(10), 1962–1973. https://doi.org/10.1037/dev0001396

Allen, R. J., Hill, L. J. B., Eddy, L. H., & Waterman, A. H. (2019). Exploring the effects of demonstration and enactment in facilitating recall of instructions in working memory. *Memory & Cognition*, *48*(3), 400–410. https://doi.org/10.3758/s13421-019-00978-6

Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology*, *106*(1), 20–29. https://doi.org/10.1016/j.jecp.2009.11.003

Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and visuospatial Short-Term and Working Memory in children: Are they separable? Child Development, 77(6), 1698–1716. https://doi.org/10.1111/j.1467-8624.2006.00968.x

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders*. https://doi.org/10.1176/appi.books.9780890425596

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), The psychology of learning and motivation: advances in research and theory (Vol. 8, pp. 47–89). New York: Academic.

Baddeley, A., Hitch, G. J., & Allen, R. (2020). A multicomponent model of working memory. In *Oxford University Press eBooks* (pp. 10–43). https://doi.org/10.1093/oso/9780198842286.003.0002

Baker-Ward, L. E., Bjorklund, D. F., & Coffman, J. (2021). The development of children's memory : the scientific contributions of Peter A. Ornstein. Cambridge University Press. https://doi.org/10.1017/9781108871105

Ben-Arieh, J., & Miller, H. J. (2009). Choosing the Interventions. In *The Educator's Guide to Teaching Students with Autism Spectrum Disorders* (pp. 55–120). Corwin Press. https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=1046409

Cassidy, A. (2013). Autism Behavior Checklist. In Springer eBooks (pp. 342–343). https://doi.org/10.1007/978-1-4419-1698-3_1367

Cohen, J. (2013). Statistical Power Analysis for the Behavioral Sciences. Academic Press.

Colley, M. (2006). Living with dyspraxia : a guide for adults with developmental dyspraxia (Rev. ed). Jessica Kingsley Publishers. https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=180445

Cook, S. W., Duffy, R. G., & Fenn, K. M. (2013). Consolidation and transfer of learning after observing handgesture. *Child Development,* 84(6), 1863–1871. https://doi.org/10.1111/cdev.12097.

Cowan, N. (2010). The Magical Mystery Four. *Current Directions in Psychological Science*, *19*(1), 51–57. https://doi.org/10.1177/0963721409359277

De Boer, A., Pijl, S. J., & Minnaert, A. (2010). Attitudes of parents towards inclusive education: a review of the literature. *European Journal of Special Needs Education*, *25*(2), 165–181. https://doi.org/10.1080/08856251003658694

Eaves, R. C., & Williams, T. O., Jr. (2006). The reliability and construct validity of ratings for the autism behavior checklist. Psychology in the Schools, 43(2), 129–142. https://doi.org/10.1002/pits.20122

Egberink, I. J. L., & Leng, W. E. (2005). COTAN Documentatie Peabody Picture Vocabulary Test. COTAN Beoordeling. Retrieved November 10, 2024, from https://www-cotandocumentatie-nl.proxy-ub.rug.nl/beoordelingen/b/13623/peabody-picture-vocabulary-test-iii-nl/

Egberink, I. J. L., & Leng, W. E. (2014). COTAN Documentatie Raven SPM verkorte versie. COTAN Beoordeling. Retrieved November 10, 2024, from https://www-cotandocumentatie-nl.proxy-ub.rug.nl/beoordelingen/b/14777/raven-spm-verkorte-versie/

Evers, A., Lucassen, W., Meijer, R., & Sijtsma, K. (2010). COTAN Beoordelingssysteem voor de kwaliteit van tests. psynip.nl. Retrieved November 30, 2024, from https://psynip.nl/wp-content/uploads/2021/11/COTAN-Beoordelingssysteem-2010.pdf

Feng, C., Wang, H., Lu, N., Chen, T., He, H., Lu, Y., & Tu, X. M. (2014). Log-transformation and its implications for data analysis. PubMed. https://doi.org/10.3969/j.issn.1002-0829.2014.02.009

Field, A. (2017). Discovering statistics using IBM SPSS statistics. https://dl.acm.org/citation.cfm?id=2502692

Garfin, D. G., & McCallon, D. (1988). Validity and reliability of the Childhood Autism Rating Scale with autistic adolescents. Journal of Autism and Developmental Disorders, 18(3), 367–378. https://doi.org/10.1007/bf02212193

Hartley, T. T., & Hitch, G. J. (2022). Working memory. *Oxford Research Encyclopedia of Psychology*. https://doi.org/10.1093/acrefore/9780190236557.013.768

Henry, L. A. (2010). The episodic buffer in children with intellectual disabilities: An exploratory study. *Research in Developmental Disabilities*, *31*(6), 1609–1614. https://doi.org/10.1016/j.ridd.2010.04.025

Hicks, S. C., Rivera, C. J., & Wood, C. L. (2015). Using direct instruction: teaching preposition use to students with intellectual disability. Language Speech and Hearing Services in Schools, 46(3), 194–206. https://doi.org/10.1044/2015_lshss-14-0088

Hollander, E., Hagerman, R. J., & Fein, D. (2018). Autism spectrum disorders. American Psychiatric Pub. https://ebookcentral.proquest.com/lib/rug/detail.action?docID=5600574

Hronis, A., Roberts, L., & Kneebone, I. I. (2017). A review of cognitive impairments in children with intellectual disabilities: Implications for cognitive behaviour therapy. *British Journal of Clinical Psychology*, *56*(2), 189–207. https://doi.org/10.1111/bjc.12133

Jaroslawska, A., Gathercole, S. E., Allen, R. J., & Holmes, J. (2016). Following instructions from working memory: Why does action at encoding and recall help? *Memory & Cognition*, *44*(8), 1183–1191. https://doi.org/10.3758/s13421-016-0636-5

JASP Team. (2024). JASP (Version 0.19.0)[Computer software]. Retrieved from https://jasp-stats.org/

Kim, S. A., & Kasari, C. (2023). Working memory of school-aged children on the autism spectrum: Predictors for longitudinal growth. *Autism*, 136236132311655. https://doi.org/10.1177/13623613231165599

Kirk, R. (2013). Experimental Design: Procedures for the Behavioral Sciences. https://doi.org/10.4135/9781483384733

Klefbeck, K. (2021). Lesson study as a way of improving school-day navigation for pupils with severe intellectual disability and autism. International Journal for Lesson and Learning Studies, 10(4), 348–361. https://doi.org/10.1108/ijlls-03-2021-0024

Luciana, M., Conklin, H. M., Hooper, C. J., & Yarger, R. S. (2005). The development of nonverbal working memory and executive control processes in adolescents. Child Development, 76(3), 697–712. https://doi.org/10.1111/j.1467-8624.2005.00872.x

Lüddeckens, J., Anderson, L., & Östlund, D. (2021). Principals' perspectives of inclusive education involving students with autism spectrum conditions – a Swedish case study. Journal of Educational Administration, 60(2), 207–221. https://doi.org/10.1108/jea-02-2021-0022norm-het-college-geeft-advies

Macrine, S. L., & Fugate, J. M. B. (2022). Embodied cognition and its educational significance. In *The MIT Press eBooks* (pp. 13–24). https://doi.org/10.7551/mitpress/13593.003.0006

Maras, K. L., Memon, A., Lambrechts, A., & Bowler, D. M. (2012). Recall of a live and personally experienced eyewitness event by adults with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *43*(8), 1798–1810. https://doi.org/10.1007/s10803-012-1729-z

Martin, L. C. (2009). The program. In *Strategies for Teaching Students With Learning Disabilities* (pp. 25-36). SAGE Publications. https://public.ebookcentral.proquest.com/choice/publicfullrecord.aspx?p=1367765

Mavilidi, M., Okely, A. D., Chandler, P., Cliff, D. P., & Paas, F. (2015). Effects of integrated physical exercises and gestures on preschool children's foreign language vocabulary learning. *Educational Psychology Review*, *27*(3), 413–426. https://doi.org/10.1007/s10648-015-9337-z

Mavilidi, M., Ouwehand, K., Okely, A. D., Chandler, P., & Paas, F. (2019). Embodying learning through physical activity and gestures in preschool children. In *Advances in Cognitive Load Theory* (pp. 103-118). Routledge.

Ministerie van Justitie en Veiligheid. (2023). Inclusief onderwijs de norm? Het college geeft advies. Nieuwsbericht | College voor de Rechten van de Mens. https://www.mensenrechten.nl/actueel/nieuws/2023/04/5/inclusief-onderwijs-de-norm-het-college-geeft-advies

Norouzian, R., & Plonsky, L. (2017). Eta- and partial eta-squared in L2 research: A cautionary review and guide to more appropriate usage. Second Language Research, 34(2), 257–271. https://doi.org/10.1177/0267658316684904

Patrick, A. (2020). Memory, Processing and SpLDs/SLDs. In *The Memory and Processing Guide for Neurodiverse Learners: Strategies for Success* (pp. 19–39). Jessica Kingsley Publishers. https://web-p-ebscohost-com.proxy-

ub.rug.nl/ehost/ebookviewer/ebook/bmxlYmtfXzIzNTY4OTBfX0FO0?sid=3b266cf6-cf52-4583-b982-59be72271f76@redis&vid=0&format=EB&rid=1

Peijnenborgh, J. C. A. W., Hurks, P. M., Aldenkamp, A. P., Vles, J. S. H., & Hendriksen, J. G. M. (2016). Efficacy of working memory training in children and adolescents with learning disabilities: A review study and meta-analysis. Neuropsychological Rehabilitation, 26(5–6), 645–672. https://doi-org.proxy-ub.rug.nl/10.1080/09602011.2015.1026356

Peng, P., & Fuchs, D. (2014). A Meta-Analysis of working memory Deficits in children with learning Difficulties. *Journal of Learning Disabilities*, *49*(1), 3–20. https://doi.org/10.1177/0022219414521667

Prelock, P., & Hutchins, T. (2018). Clinical Guide to Assessment and Treatment of Communication Disorders. In *Best practices in child and adolescent behavioral health care*. https://doi.org/10.1007/978-3-319-93203-3

Raven, J., Court, J. H., & Raven, J. (1998). *Manual for Raven's progressive matrices and vocabulary scales*. Oxford Psychologists Press.

Richardson, J. T. (2011). Eta squared and partial eta squared as measures of effect size in educational research. Educational Research Review, 6(2), 135–147. https://doi.org/10.1016/j.edurev.2010.12.001

Ring, M., Guillery-Girard, B., Quinette, P., Gaigg, S. B., & Bowler, D. M. (2020). Short-Term memory span and Cross-Modality integration in younger and older adults with and without autism spectrum disorder. *Autism Research*, *13*(11), 1970–1984. https://doi.org/10.1002/aur.2387

Rivet, T. T., & Matson, J. L. (2011). Gender Differences in Core Symptomatology in Autism Spectrum Disorders across the Lifespan. Journal of Developmental and Physical Disabilities, 23(5), 399–420. https://doi.org/10.1007/s10882-011-9235-3

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.

Schalock, R. L., Luckasson, R., & Tassé, M. J. (2021). An Overview of Intellectual Disability: Definition, Diagnosis, Classification, and Systems of Supports (12th ed.). American Journal on Intellectual and Developmental Disabilities, 126(6), 439–442. https://doi.org/10.1352/1944-7558-126.6.439

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

Sepp, S., Howard, S. J., Tindall-Ford, S., Agostinho, S., & Paas, F. (2019). Cognitive Load Theory and Human Movement: Towards an Integrated Model of Working Memory. Educational Psychology Review, 31(2), 293–317. https://doi.org/10.1007/s10648-019-09461-9

Shurr, J., Minuk, A., Holmqvist, M., Östlund, D., Ghaith, N., & Reed, B. (2021). Parent perspectives on inclusive education for students with intellectual disability: A scoping review of the literature. International Journal of Developmental Disabilities, 69(5), 633–643. https://doi.org/10.1080/20473869.2021.2003612

Thurm, A., Farmer, C., Salzman, E., Lord, C., & Bishop, S. (2019). State of the field: Differentiating intellectual disability from autism spectrum Disorder. Frontiers in Psychiatry, 10. https://doi.org/10.3389/fpsyt.2019.00526

Tiernan, B., Casserly, A. M., & Maguire, G. (2018). Towards inclusive education: instructional practices to meet the needs of pupils with special educational needs in multi-grade settings. International Journal of Inclusive Education, 24(7), 787–807. https://doi.org/10.1080/13603116.2018.1483438

Tindall-Ford, S., Agostinho, S., & Sweller, J. (2020). Advances in cognitive load theory : rethinking teaching. Routledge. https://doi.org/10.4324/9780429283895

Tryfon, M., Anastasia, A., & Eleni, R. (2019). Parental perspectives on inclusive education for children with intellectual disabilities in Greece. International Journal of Developmental Disabilities, 67(6), 397–405. https://doi.org/10.1080/20473869.2019.1675429

Van Buuren, S. (2018). Flexible Imputation of Missing Data, second edition. In Chapman and Hall/CRC eBooks. https://doi.org/10.1201/9780429492259

Wang, L., Xie, T., Ma, H., Xu, M., & Xie, X. (2022). Subject-performed task effect on working memory performance in children with autism spectrum disorder. Autism Research : Official Journal of the International Society for Autism Research, 15(9), 1698–1709. 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. (2022). 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. https://doi.org/10.1007/s10803-022-05780-z

Xie, T., Wang, L., & Wang, T. (2021). In which case is working memory for movements affected by verbal interference? Evidence from the verbal description of movement. *Memory*, *29*(6), 762–777. https://doi.org/10.1080/09658211.2021.1944217

Yang, T., Gathercole, S. E., & Allen, R. J. (2013). Benefit of enactment over oral repetition of verbal instruction does not require additional working memory during encoding. Psychonomic Bulletin & Review, 21(1), 186–192. https://doi.org/10.3758/s13423-013-0471-7

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

Footnotes

[1] The WM consists of four parts with their own function. Firstly, the information enters the memory in the central executive which orders information. Secondly, visual information goes to the visuospatial sketchpad and verbal information goes into the phonological loop. Finally, the episodic buffer integrates information form the visuospatial sketchpad and the phonological loop into a coherent whole. (Baddeley and Hitch, 1974)

Appendix A

Assumptions coded contrast repeated measures ANOVA

Below the assumption checks for the coded contrast repeated measures ANOVA are described

Independent observations

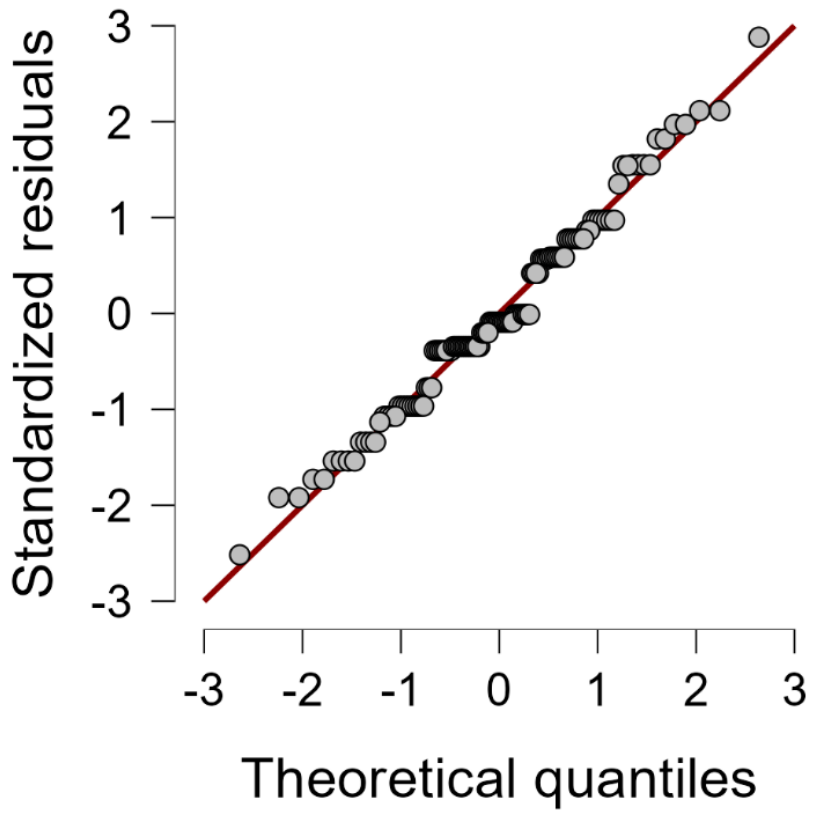
As described in the method, the sample that was drawn for this study was random, therefore this assumptions has been met

Normality

The assumption of normality was checked via the Q-Q Plot in Figure A1 As the findings are caught by a linear line the assumption of normality has been met. However, when the Shapiro-Wilk test was conducted (Table A1), not all groups were significant. In the ID group in the enactment condition p=0.318 and the TD group in the hearing condition p=0.051. To correct for the non-normality in these groups both the enactment and the hearing variable log transformed and the tests were conducted again. In Table A2 and Figure A2 the

Figure A1

Q-Q plot of the residuals



Note. Analysis performed on data from "*Can Enactment and Motor Imagery Improve Working Memory for Instructions in Children with Autism Spectrum Disorder and Children with Intellectual Disability?"* by Xie et al. (2022) *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-022-05780-z

results are shown. The Shapiro-Wilk outcomes for both problematic groups, ID in the enactment (p=0.137) condition and TD in the hearing condition (p=0.068), are still not significant. In addition, the Q-Q plot in A2 now shows an

Table A1

Shapiro-Wilk test in the hearing and enactment condition

	Hearing			Enactment		
	ASD	ID	TD	ASD	ID	TD
Shapiro-Wilk	0.869	0.798	0.905	0.694	0.947	0.888
p-value	0.011	<.001	0.051	<.001	0.318	0.024

Note. Analysis performed on data from "*Can Enactment and Motor Imagery Improve Working Memory for Instructions in Children with Autism Spectrum Disorder and Children with Intellectual Disability?"* by Xie et al. (2022) *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-022-05780-z

outlier. Therefore, in this case the log-transformation does not seem fitting and the original data will be used. It is important to note that all outcomes should be handled with caution because the assumption of normality was not met completely.

Table A2

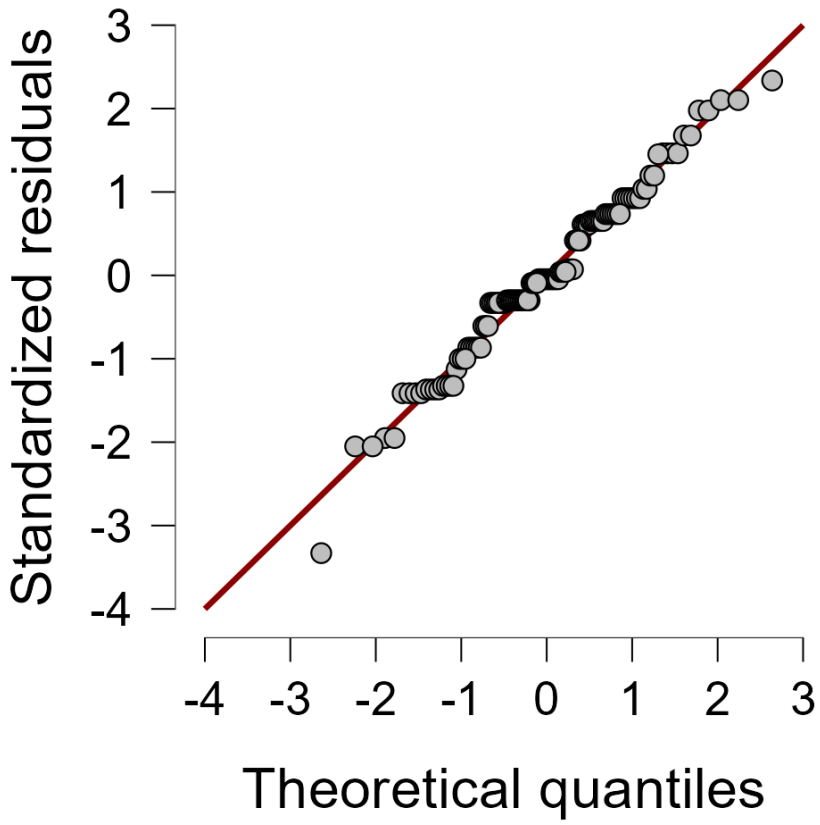
Shapiro-Wilk test on the log transformed scores in hearing and enactment

	Hearing			Enactment		
	ASD	ID	TD	ASD	ID	TD
Shapiro-Wilk	0.875	0.816	0.912	0.721	0.927	0.877
p-value	0.014	0.001	0.068	<.001	0.137	0.015

Note. Analysis performed on data from "*Can Enactment and Motor Imagery Improve Working Memory for Instructions in Children with Autism Spectrum Disorder and Children with Intellectual Disability?"* by Xie et al. (2022) *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-022-05780-z

Figure A2

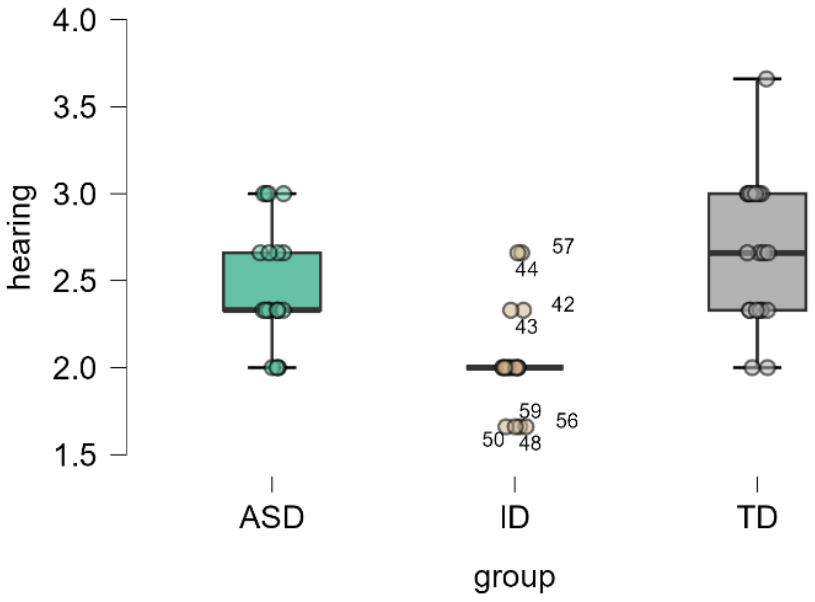
Q-Q plot of the residuals after log-transformation



Note. Analysis performed on data from "*Can Enactment and Motor Imagery Improve Working Memory for Instructions in Children with Autism Spectrum Disorder and Children with Intellectual Disability?"* by Xie et al. (2022) *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-022-05780-z

Figure A3

Boxplots of the distribution of span scores in the hearing condition in all groups

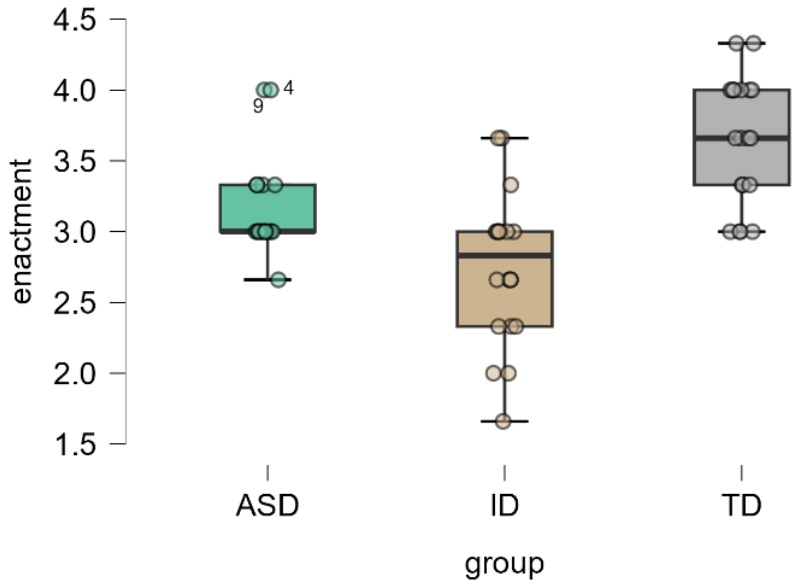


Note. Analysis performed on data from "*Can Enactment and Motor Imagery Improve Working Memory for Instructions in Children with Autism Spectrum Disorder and Children with Intellectual Disability?"* by Xie et al. (2022) *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-022-05780-z

Outliers

Figure A4

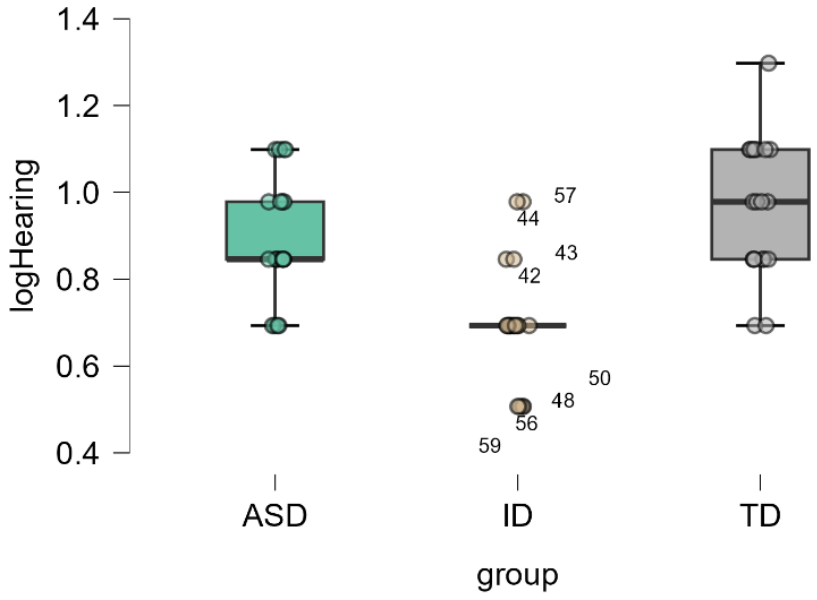
Boxplots of the distribution of span scores in the enactment condition in all groups



Note. Analysis performed on data from "*Can Enactment and Motor Imagery Improve Working Memory for Instructions in Children with Autism Spectrum Disorder and Children with Intellectual Disability?"* by Xie et al. (2022) *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-022-05780-z

Figure A5

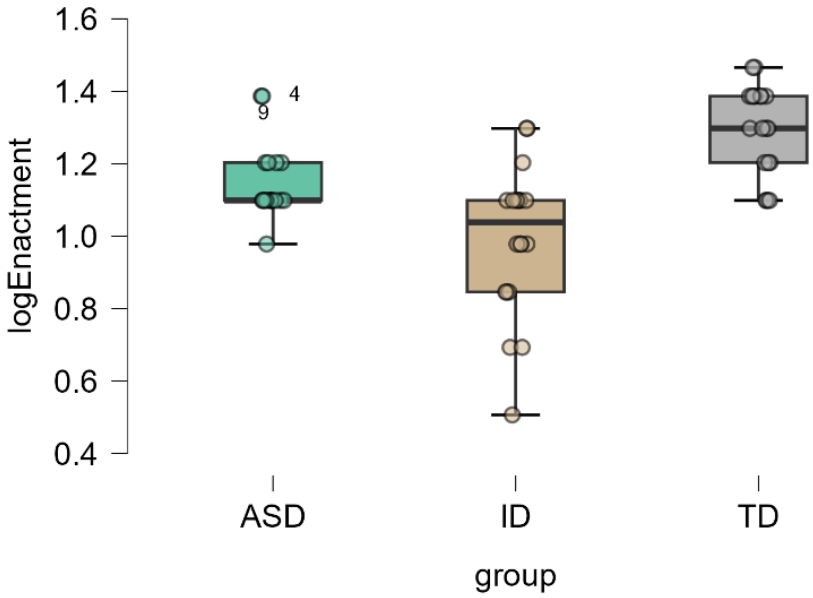
Boxplots of the distribution of the log transformed span scores in the hearing condition in all groups



Note. Analysis performed on data from "*Can Enactment and Motor Imagery Improve Working Memory for Instructions in Children with Autism Spectrum Disorder and Children with Intellectual Disability?"* by Xie et al. (2022) *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-022-05780-z

Figure A6

Boxplots of the distribution of the log transformed span scores in the enactment condition in all groups



Note. Analysis performed on data from "*Can Enactment and Motor Imagery Improve Working Memory for Instructions in Children with Autism Spectrum Disorder and Children with Intellectual Disability?"* by Xie et al. (2022) *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-022-05780-z

To test for influential outliers boxplots where made. The boxplots in Figure A3 and Figure A4 is show that in both the ID group in the hearing condition as the ASD group in the enactment condition there are influential outliers. Therefore, a log-transformation of the hearing and the enactment variable was made. With these two log-variables new boxplots were made (Figure A5 and Figure A6). As visible, the log-transformation did not help with the outliers. Therefore, in this case the log-transformation does not seem fitting and the original data will be used. It is important to note that all outcomes should be handled with caution because the assumption of normality was not met completely.

Equal variances

The assumption for equal variances was tested with the Levene's test (Table A3), for both conditions (hearing and enactment) the score was not significant (hearing: p= 0.078, enactment: p=0.096). This is proof that the assumption of equal variances has been met.

Table A3

Levene's test for equal variances groups

	F	df1	df2	p
Hearing	2.669	2	57	0.078
Enactment	2.444	2	57	0.078

Note. Analysis performed on data from "*Can Enactment and Motor Imagery Improve Working Memory for Instructions in Children with Autism Spectrum Disorder and Children with Intellectual Disability?"* by Xie et al. (2022) *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-022-05780-z

Appendix B

Assumptions ANCOVA

Below the different assumptions for ANCOVA are explained

Independent observations

The first assumption that has to be met, before an ANCOVA can be conducted, is independent observations, this is discussed in the method section.

Homogeneity of regression slopes

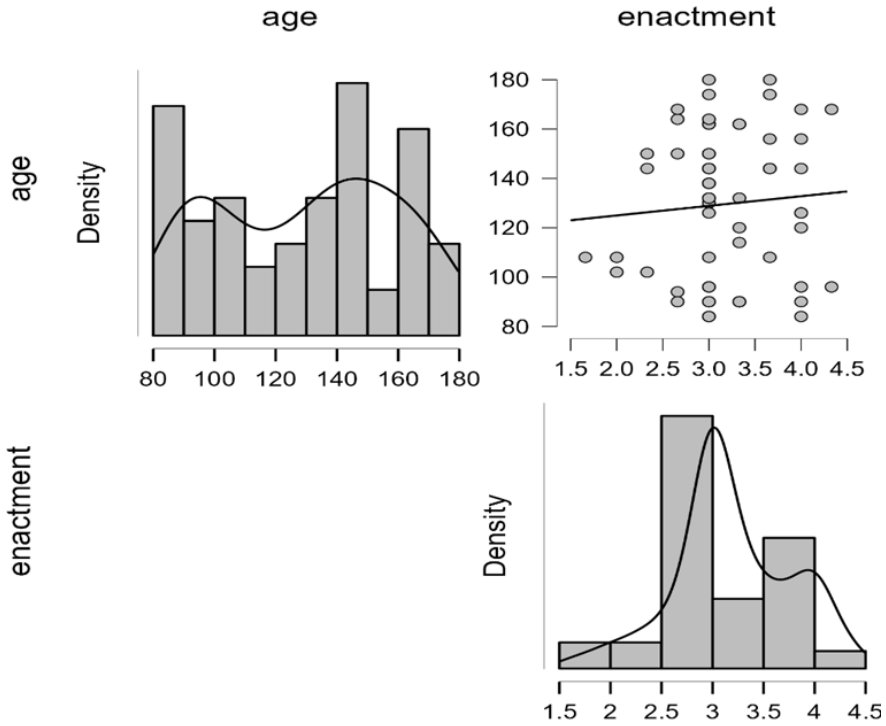
Second there needs to be homogeneity of regression slopes. To test for this assumption the interaction between group and age was plotted. The interaction effect was not significant (F=0.063, p=0.803), therefore there are no signs of interaction and the assumption has been met.

Linearity between the covariate and the dependent variable

The assumption of linearity between the covariate and the dependent variable was tested by making a scatter plot with age on the y-axis and span score in the enactment condition on the x-axis. As shown in the scatterplot Figure B1 the linear relationship between age and enactment is weak. Therefore the assumption of linearity is violated and is it not possible to conduct an ANCOVA.

Figure B1

Descriptive plots – age and enactment



Note. Analysis performed on data from "*Can Enactment and Motor Imagery Improve Working Memory for Instructions in Children with Autism Spectrum Disorder and Children with Intellectual Disability?"* by Xie et al. (2022) *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-022-05780-z

Equal variances

To test for the assumption of equal variances a Levene's test was conducted. As shown in Table B1 the test was not significant (p=0.105) therefore the assumption of equal variances has been met.

Table B1

Levene's test for equal variances

F	*df1*	*df2*	*p*
2.342	2.000	57.000	0.105

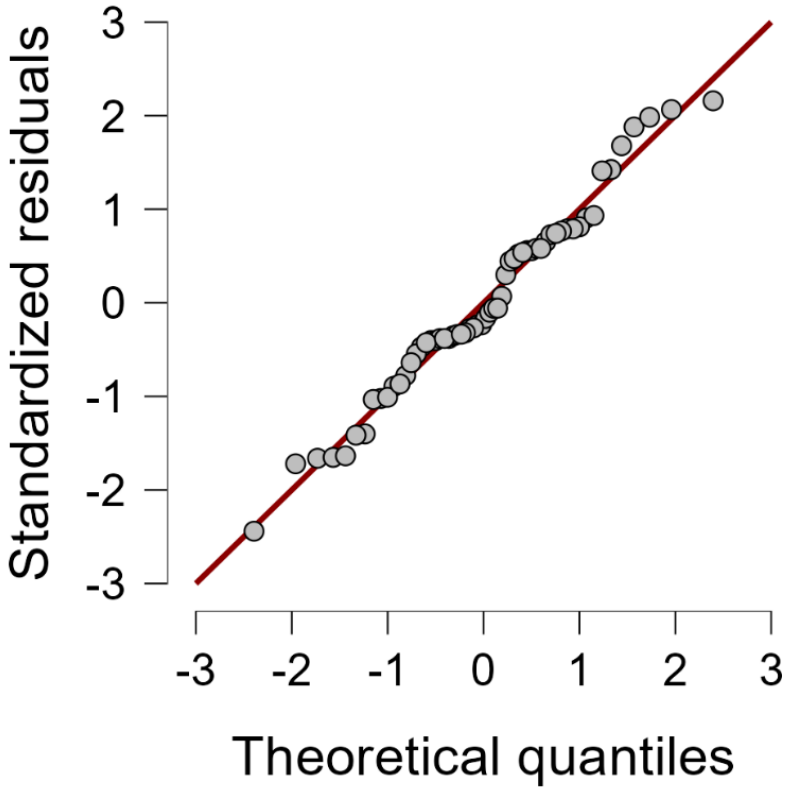
Note. Analysis performed on data from "*Can Enactment and Motor Imagery Improve Working Memory for Instructions in Children with Autism Spectrum Disorder and Children with Intellectual Disability?"* by Xie et al. (2022) *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-022-05780-z

Normality

To test for the assumption of normality of the residuals a Q-Q plot was made. When there is a linear relation the residuals are normally distributed. As shown in Figure B2 there is a linear relation therefore the assumption has been met.

Figure B2

Q-Q plot of the residuals



Note. Analysis performed on data from "*Can Enactment and Motor Imagery Improve Working Memory for Instructions in Children with Autism Spectrum Disorder and Children with Intellectual Disability?"* by Xie et al. (2022) *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-022-05780-z

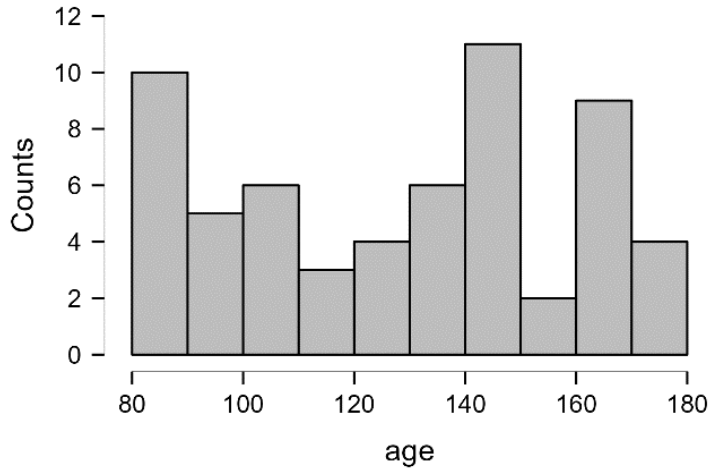
Appendix C

Assumptions Spearman correlation

Because the linear relation between the covariate and the dependent variable was very weak and the distribution of both variables was not normal Figure C1 and Figure C2, a spearman correlation was conducted. The assumptions for the Spearman correlation are shown below.

Figure C1

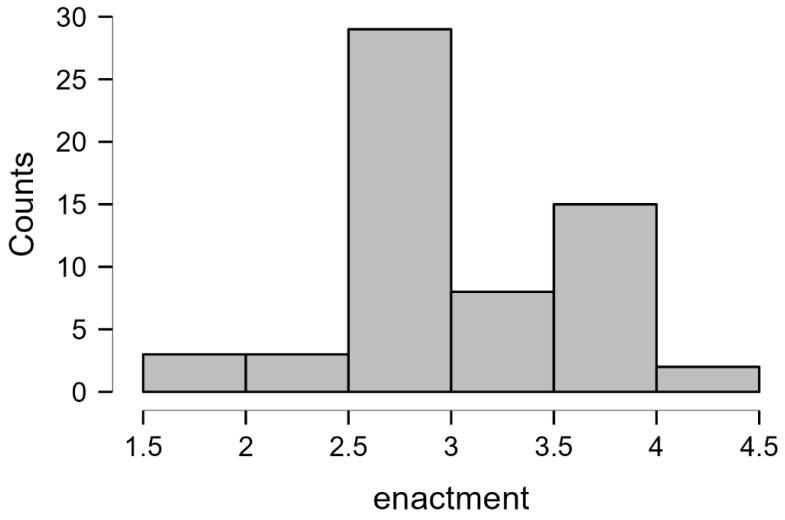
Distribution of age across all groups



Note. Analysis performed on data from "*Can Enactment and Motor Imagery Improve Working Memory for Instructions in Children with Autism Spectrum Disorder and Children with Intellectual Disability?"* by Xie et al. (2022) *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-022-05780-z

Figure C2

Distribution of span scores in the enactment condition across all groups



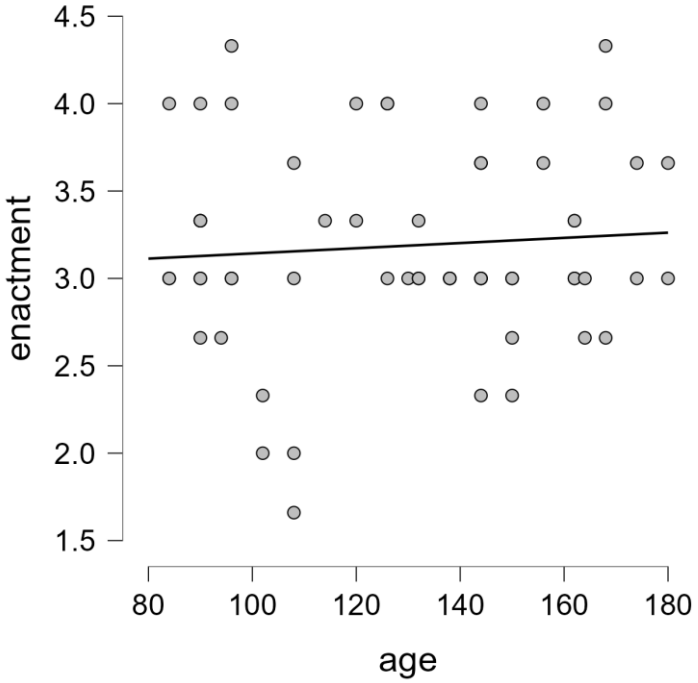
Note. Analysis performed on data from "*Can Enactment and Motor Imagery Improve Working Memory for Instructions in Children with Autism Spectrum Disorder and Children with Intellectual Disability?"* by Xie et al. (2022) *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-022-05780-z

Monotonic relationship

The scatterplot in Figure C3 shows that the data has a weak linear relation, with that the data is monotonic

Figure C2

Distribution of span scores in the enactment condition across all groups



Note. Analysis performed on data from "*Can Enactment and Motor Imagery Improve Working Memory for Instructions in Children with Autism Spectrum Disorder and Children with Intellectual Disability?"* by Xie et al. (2022) *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-022-05780-z