Psicologia: Teoria e Pesquisa
2025, v.41, e41201

DOI: https://doi.org/10.1590/0102.3772e41201.en

Behavioral Sciences

Assessment of Working Memory in Students with Autism and/or Intellectual Disabilities Using Eye Tracking

Heloísa D'Angelo[1], Jordão Frazão[2], Tardelly Araújo Cavalcante[2],
Priscila Benitez[1], Camila Domeniconi[3] & André Castelo Branco Soares[2]

[1]*Universidade Federal do ABC, Santo André, SP, Brasil*

[2]*Universidade Federal do Piauí, Teresina, PI, Brasil*

[3]*Universidade Federal de São Carlos, São Carlos, SP, Brasil*

ABSTRACT – Working memory was assessed with delayed matching-to-sample, with eye tracking. The case study aimed to analyze eye tracking in a matching-to-sample task with delays of 0, 4 and 8 seconds, and 20 minutes, with social and non-social stimuli. Three students participated. E1 (autism, 6 years old) had a statistical difference in the 8-second delay. E2 (autism and ID, 12 years old) and E3 (ID/Down Syndrome, 17 years old) showed significant difference between social and non-social stimuli (E2=20.5% for non-social and 2.83% for social; E3=63.01% for non-social and 1.23% for social). The data are expected to assist in the planning of teaching procedures.

KEYWORDS: autism, intellectual disability, eye tracking, working memory

Avaliação da Memória de Trabalho de Estudantes com Autismo e/ou Deficiência Intelectual Por Meio do Rastreamento Ocular

RESUMO – A memória de trabalho foi avaliada com emparelhamento com o modelo com atraso, com rastreio ocular. O estudo de caso objetivou analisar o rastreamento ocular em tarefa de emparelhamento com o modelo com atrasos de 0, 4 e 8 segundos, e 20 minutos, com estímulos sociais e não sociais. Participaram três estudantes. E1 (autismo, 6 anos) teve diferença estatística no atraso de 8 segundos. E2 (autismo e DI, 12 anos) e E3 (DI/Síndrome de Down, 17 anos) apresentaram diferença significativa entre estímulos sociais e não sociais (E2=20,5% para não sociais e 2,83% para sociais; E3=63,01% para não sociais e 1,23% para sociais). Espera-se que os dados auxiliem no planejamento de procedimentos de ensino.

PALAVRAS-CHAVE: autismo, deficiência intelectual, memória de trabalho, rastreamento ocular

The study of cognitive development can assist in understanding the functioning of executive functions, particularly those related to working memory, and their impact on the schooling process. Previous literature (Czermainski *et al.*, 2013; Dias & Seabra, 2013; Gaviria & Fitzgerald, 2014; Maranhão & Pires, 2017; Siquara *et al.*, 2014) has shown the importance of working memory in maintaining information in a temporary time interval and the ability to manage information to execute specific tasks or to keep track of updates in an activity.

Planning and organizing sequences of actions in the future are complex behaviors that can be learned throughout the schooling process. This means that working memory refers to a process that requires the ability to learn (Siquara *et al.*, 2014). Therefore, evaluating and teaching working memory can contribute to student retention as well as prevent school dropout, as already indicated in previous literature (Dias & Seabra, 2013).

In the context of Special Education, from the perspective of inclusive education, students with autism and/or intellectual disabilities (ID) can benefit from teaching procedures aimed at stimulating working memory. As documented in a previous study (Czermainski *et al.*, 2013), students with autism and/or ID may present deficits related to working memory, thus necessitating procedures to ensure the teaching of executive functions (Dias & Seabra, 2013).

Autism is referred to as Autism Spectrum Disorder (APA, 2013) and causes persistent deficits in communication and social interaction in various contexts, stemming from alterations in socioemotional reciprocity, communicative behaviors, and interpersonal relationships. In the United States, one in every 54 children aged eight or younger is diagnosed with ASD (Maenner *et al.*, 2016). In developed countries, the prevalence varies from 0.67% to 1.13%, being four times more common in males than females (Maenner *et al.*, 2016).

In Brazil, there is still no precise estimate of the prevalence of ASD. A pilot study, focusing on the São Paulo region, identified that 0.3% of Brazilians may have ASD, meaning that approximately 40,000 children or adolescents (up to 20 years old) could have the disorder (Paula *et al.*, 2011). The rise in the number of ASD diagnoses is due to various factors, ranging from increased awareness to more sophisticated research, which defines the critical components of the disorder more accurately (Altenmuller-Lewis, 2017).

As it is a spectrum, symptoms vary from mild to moderate to severe, and according to DSM-5 (APA, 2013), intervention should occur based on the type of support needed for each case. The heterogeneity of symptoms leads to behavioral variability, meaning that some people with ASD may, for example, develop fluent speech while others may remain nonverbal (Coderre *et al.*, 2019).

A critical variable for the prognosis of ASD concerns the presence or absence of intellectual disability (ID). ID is common in ASD (APA, 2013). The causes of ID in ASD are multifaceted, ranging from the intensity of intervention hours to the age of intervention onset, among others (APA, 2013). ID can be defined as a disorder that begins in the developmental period, characterized by deficits in intellectual and adaptive functions, with impairments in areas such as reasoning, problem-solving, experiential learning, personal independence, and social responsibility, among other domains. The disorder is categorized by levels (mild, moderate, severe, and profound), classified based on intellectual and adaptive functioning expressed in social, conceptual, and practical skills (APA, 2013). Both ID and ASD are referred to in DSM-5 (APA, 2013) as neurodevelopmental disorders.

Dawson and colleagues found in 2002 that children with ASD showed no difference in evoked potentials between the presentation of a familiar or unfamiliar face but did so when presented with a familiar and a new object. Children without ASD had significant differences in potential for both objects and faces. In the study by Hauck *et al.* (1998), impairment in the performance of children with ASD was found in a face memory task compared to children without the disorder, but not in face matching tasks, nor in memory or object matching tasks.

Understanding that the behavior of remembering (in this case, memory tasks) can be studied through the programming of successive delays, Teixeira (2019) investigated the effect of delay times on conditional, identity, and arbitrary discrimination tasks with adolescents with and without ID. The programmed delays were 0, 2, 4, 6, and 8 seconds. The results showed that with increasing delay, there was a decrease in the performance of participants, mainly those with ID, specifically participants with Down syndrome. Four out of six participants with ID required additional procedures to establish arbitrary relationships between stimuli.

One way to investigate other variables in tasks involving delay, in addition to measures of accuracy and error, is using eye tracking, a technique that has become increasingly common in studies of this nature. It is a technology that maps visual behavior, based on the detection of different parameters in real-time, such as the route of the gaze, saccadic movements, and fixation. The equipment records how many times (fixation) and for how long (in milliseconds) (fixation duration) the student looks at the stimuli presented in a computer task given. The operation occurs as follows: an infrared beam is projected onto the eyes of the person being evaluated, which causes a reflection in the pupil that is then captured by a sensor capable of identifying the direction from which the gaze comes and measuring the parameters of average duration, length, and location (Orsati et al, 2009; Schwartzman *et al.*, 2015; Zangrando, 2018).

Scientific interest in the use of eye tracking has increased substantially in recent years, as it allows for an objective evaluation of ocular behavior during tasks (Zangrando, 2018), especially with Brazilian students with ASD (Schwartzman *et al.*, 2015), Rett Syndrome (Schwartzman *et al.*, 2015), Global Developmental Disorder (Orsati *et al.*, 2009), and comparisons of students with and without disorders (Schwartzman *et al.*, 2015). Interest has varied in terms of identifying patterns of ocular movement during the presentation of diverse stimuli (Zangrando, 2018). Stimuli can be grouped into social (emotional, familiar faces, unfamiliar faces, inverted) and non-social (objects).

One aspect that has been investigated concerns both the interest of participants with ASD in social stimuli (essential for the development of empathy, for example) and their eye tracking. In the case of discriminating emotions in human faces, participants without ASD exhibit eye fixations in the T area (eye-eye-nose-mouth regions), while people with ASD exhibit diffuse fixations and fixation times in areas peripheral to the T area (Muñoz, 2018; Orsati *et al.*, 2009; Zangrando, 2018).

Experimental tasks using model matching have been used to investigate how the construction of relationships between stimuli occurs for students with different profiles of learning and development. The model matching task can be organized as a succession of discrete trials, in which the student is instructed to choose a comparison stimulus, depending on the model stimulus. In a model matching discrimination trial, the model stimulus and comparison stimuli can be presented simultaneously (simultaneous matching), or the comparison stimuli can be presented after the disappearance of the model stimulus (delayed model matching, often referred to delayed matching-to-sample). This is a critical variable in the format and organization of the task that can impact the results obtained, with existing findings indicating the use of delay matching tasks as conducive to more robust learning of stimulus relationships (for example, Bortoloti & de Rose, 2009; 2012).

Literature has found that individuals with ASD with or without ID present global deficits in executive functioning, with working memory highlighted as the area with the greatest deficit evaluated in the study by McClain *et al.* (2022). The study by Roberts and Richmond (2014) identified specific deficits in learning and memory processes in individuals with Down Syndrome, due to ID.

Considering the cited studies, the originality of this study refers to the combination of two variables: a) programming of delay time in seconds and minutes; and b) type of stimulus - social and non-social. Thus, the question arises as to whether a smaller/greater delay time results in a smaller/ greater number of correct responses in both types of stimuli.

Additionally, the performance of a student with ASD, another with ASD and ID, and another with ID/Down Syndrome is also investigated. Therefore, this study aimed to analyze the pattern of eye tracking during delay tasks (0, 4, and 8 seconds and 20 minutes), using social stimuli (faces) and non-social stimuli (objects) in students with ASD and/or ID. One way to assess and teach working memory in a more elementary manner may be with delayed model matching procedures. The present study proposes to analyze eye tracking in delayed model matching tasks (0, 4, and 8 seconds and 20 minutes), using social (faces) and non-social (objects) stimuli in students with ASD and/or ID.

MATERIALS AND METHODS

Participants

Three students participated, two with a medical diagnosis of ASD and one with Down syndrome and intellectual disability (ID). The selection criteria for the study were a medical diagnosis of ASD and/or ID, undergoing behavioral intervention, and being under 18 years of age. Table 1 shows the description of the participants' characteristics.

The inclusion criteria for the study were the presentation of a medical report by the mother of the student. The information in Table 1 was based on the supplied report. All three students were verbal, with chronological ages ranging from 6 to 17 years. All three were enrolled in regular school in mainstream classrooms, lacking reading, writing, and math skills. Only E1 was undergoing intensive behavioral intervention. E2 and E3 engaged in activities provided by their respective mothers, with guidance from a behavior analyst. E2 had mild ID and ASD, with significant deficits in the academic domain compared to other areas. E1 and E3, on the other hand, had moderate ID, with significant deficits in both oral and written language. E1 exhibited the lowest repertoire of social skills.

Materials and Location

A portable eye-tracking device (Tobii brand) was used to record eye movement patterns, and a touchscreen notebook was used for the task execution by the participants. Data collection took place in a meeting room at the Federal University of ABC, as well as in a specialized clinic.

Variables

The independent variable involved the application of receptive vocabulary tasks (through auditory-visual matching

Table 1. Characterization of the students

Student	Chronological age	Gender	Current grade in Brazilian school system	Diagnosis
E1	6	Male	1st grade	ASD
E2	12	Female	5th grade	ASD with ID
E3	17	Male	9th grade	ID / Down Syndrome

tasks) and expressive vocabulary tasks (naming) in both the baseline phase without delays, where each trial was presented one after the other, and the delayed phase with delays of 0s, 4s, 8s, and 20 minutes. The dependent variable included measures of the number of correct responses in auditory-visual matching (receptive vocabulary) and naming (expressive vocabulary) tasks with delays of 0, 4, and 8 seconds and 20 minutes. Another dependent variable analyzed was the time spent looking at the computer screen for each trial (gaze presence - the amount of time the sensor detected gaze for each presented trial), heat map, and fixation positions (scan path). The heat map is used to represent the time and number of eye fixations for regions of the computer screen, where a higher concentration of warm colors (such as red and yellow) indicates more time and eye fixation in a specific region.

Procedure

Phase 1 - Development of computerized tasks: Two tasks were proposed for characterizing receptive and expressive vocabularies: auditory-visual matching and naming of non-social stimuli, without eye-tracking measures. The delayed matching-to-sample task was used for evaluation with delays, employing both social and non-social stimulus groups, with eye-tracking measures. This task was chosen to obtain accuracy and error measures, as well as the number of eye fixations, based on the delay time programmed in the procedure. For the non-social stimulus group, 60 visual and auditory stimuli were randomly distributed into 30 pairs, without any specific criteria (30 were photographs of real objects, and 30 were digital illustrations - drawings).

The delayed matching-to-sample task employed to evaluate memory behavior involved the presentation of an image in the center of a blank screen. No oral instructions were provided; it was expected that the student would click on the presented stimulus. Clicking on the stimulus led to the appearance of another blank screen, with a duration determined by the delay (*e.g.*, 0s, 4s, or 8s). Subsequently, a blank screen with two stimuli was presented at the bottom of the screen, one on the left corner and the other on the right, with the target stimulus presented previously and a novel one, accompanied by the instruction: "Which one did you see before?". The student was instructed to select the stimulus, receiving social praise for correct responses and proceeding to the next trial for incorrect responses. Furthermore, additional reinforcing consequences were programmed, such as access to a cell phone for about three minutes after a correct response. These consequences were not standardized and were only used with E1 to keep him engaged in the task. Figure 1 shows the visual organization of the delayed tasks.

The stimuli used were images of everyday objects for non-social stimuli (such as clothing, toys, and kitchen utensils) and social stimuli with pictures of people. All images were sourced from the FreePik free image bank. Both the non-social and social stimulus sets consisted of 30 pairs of stimuli. A stimulus pair refers to a set of two images, from which the student had to select one. The non-social stimuli were divided into two balanced categories: photographs of real objects and digital illustrations (drawings). Three balanced categories were used for social stimuli: photographs of adults, photographs of children, and digital illustrations of faces (drawings).

Table 2 shows the organization of stimuli at each delay. It is important to note that the application was conducted similarly for all students; however, the equipment did not record the performance of student E1 in the non-social stimulus phase, and the performance of student E2 in the 20-minute delay phase was not analyzed due to a procedural error.

Phase 2 - Characterization of students' receptive (listener) and expressive (speaker) vocabulary: Phases 2, 3, and 4 were administered on the same day, one after the other. For the evaluation of receptive vocabulary, an assessment was done using the same images as the non-social stimulus task, totaling 60 stimuli. The objective was to assess each student's baseline repertoire in identifying the figure in response to the auditory stimulus, as well as naming the figures independently. In this case, two figures were displayed on the screen, followed by the instruction "Show X," where X is the name of the object. Social praise was given for correct responses, and incorrect responses led to the next trial. For the evaluation of expressive vocabulary, the student was instructed to name

Figure 1. Sequence of slides presented during delayed matching-to-sample task.

Table 2. Organization of stimuli in delayed matching tasks

Stage	#	Delay times	Types of stimuli
Nonsocial stimuli	30 pairs	10 pairs of stimuli with 0 seconds 10 pairs of stimuli with 4 seconds 10 pairs of stimuli with 10 seconds	5 pairs of photos for each category of delay times 5 pairs of digital illustrations for each category of delay times
Social Stimuli	30 pairs	10 pairs of stimuli with 0 seconds	4 pairs of photos of children (2 girls and 2 boys) 3 pairs of photos of adults (1 woman and 2 men) 3 pairs of digital illustrations of faces
		10 pairs of stimuli with 4 seconds 10 pairs of stimuli with 10 seconds	3 pairs of photos of children (1 girl and 2 boys) 3 pairs of photos of adults (2 women and 1 man) 4 pairs of digital illustrations of faces 3 pairs of photos of children (2 girls and 1 boy) 4 pairs of photos of adults (2 women and 2 men) 4 pairs of digital illustrations of faces
Social and nonsocial stimuli	12 pairs	20 minutes (in each pair, one stimulus had been presented before, and one was novel)	6 pairs of nonsocial stimuli 6 pairs of social stimuli

each figure ("What is this?"), with the same consequence programming as before. All non-social stimulus images used throughout the experiment were evaluated. This phase was administered in a single session, with two blocks presented, each consisting of 30 consecutive trials for the listening task and another with 30 naming trials.

Phase 3 - Calibration of the eye-tracking device: Calibration was performed with an image sampling rate of 60 Hz with Unity standard quality, using a six-point calibration. For this, seven blue points in the form of dots were presented on a black background on the computer screen, and the student was instructed to look at each point until it disappeared from the screen. The procedure followed the findings of Huang and Bulling (2019). After completing the calibration, the student proceeded to the delayed matching-to-sample assessment. The calibration time varied among students, with E1 taking around 40 minutes to complete this phase, E2 approximately five minutes, and E3 about 2-3 minutes.

Phase 4 - Assessment with delays: The assessment sequence used the non-social stimulus set with delays of 0s, 4s, and 8s, followed by the social stimulus set. After 20 minutes from the assessment session, 12 stimulus pairs (6 social and 6 non-social) were presented. Each correct stimulus appeared in alternate positions, randomized, to minimize sequence effects. The experimental task was consistent across this phase, involving the presentation of the target stimulus on a blank screen, followed by a touch from the student, then presenting a blank screen (depending on the delay of 0s, 4s, or 8s), followed by the presentation of two stimuli, one correct and one distractor.

The comparison stimuli presented on the screen served two different functions: target (to be indicated by the student, considered correct) or distractor (incorrect stimulus). Additionally, 12 stimuli that were not presented previously at any point in the assessment were used for the 20-minute delay. In this assessment, the previously unseen stimuli were presented to the student along with an experimental stimulus, and the student was asked to identify the one they had seen before by touching the computer screen.

Data Analysis

The data were analyzed based on a single-subject design, comparing participants' performance with themselves throughout the entire experimental exposure. The dependent variables analyzed were the number of correct responses and implicit measures such as time spent looking at the computer screen between each trial (gaze presence - the amount of time the sensor detected gaze for each presented trial), the gaze heat map, and fixation positions. Rstudio software was used to generate graphical analysis of the data captured by the eye-tracking device.

The time spent fixating on the screen among different delay times was evaluated with a one-way ANOVA test, with delay time as the independent variable, with Tukey post-hoc test for subsequent verification of differences between groups. Similarly, the time spent fixating on the screen between social and non-social stimuli was also evaluated with a t-test, with stimulus type as the independent variable.

The characterization of receptive and expressive vocabulary was analyzed in five categories: *Independent correct responses:* The student responded correctly without any assistance; *Partial verbal* cue (in expressive vocabulary): The first syllable of the word was spoken, and the participant completed the response correctly; *Total verbal cue* (in expressive vocabulary)*:* The requested word was spoken, and the participant repeated the response correctly; *Gestural cue:* A gesture resembling the function of the requested object was made (*e.g.*, a sweeping gesture for an image of a broom); *Total physical cue:* The student's hand was guided to the correct object.

Correct and incorrect responses were evaluated based on automatic verification of clicks on the computer touchscreen. A standard was established for the proximity that the click should be of each image to consider the response correct or incorrect and clicks exactly in the middle between the two images were considered incorrect responses.

RESULTS

Characterization of receptive (listening) and expressive (speaking) vocabulary

The characterization data were analyzed in terms of hits and misses for each stimulus, according to Figure 2. In the expressive vocabulary test, words with very similar meanings to the reference were considered hits ("Pot" for "Bowl," etc.). The three students showed a higher number of independent responses for receptive vocabulary compared to expressive, meaning that they could identify figures independently when orally instructed but still needed support to verbally name them.

Performance of E1, E2, and E3 in delayed matching tasks

For student E1, Table 3 shows the number of attempts evaluated for each delay time, *i.e.*, those in which gaze presence time and touch on the touchscreen could be measured. The standard deviation shows how much these data vary in relation to the mean. In this case, high variance in the time spent in saccadic movement was identified. For the analysis of time spent on saccadic movements, only delays of 0, 4, and 8 seconds were used. The standard deviation shows how much data for E2 (Table 3) vary in relation to the mean, also demonstrating high variance in time spent in saccadic movement. The t-test showed that there was a significant difference ($t = 3.763$, $p = 0.00034$) in gaze presence averages between social and non-social stimuli. The t-test was used as it is recommended for analysis between two

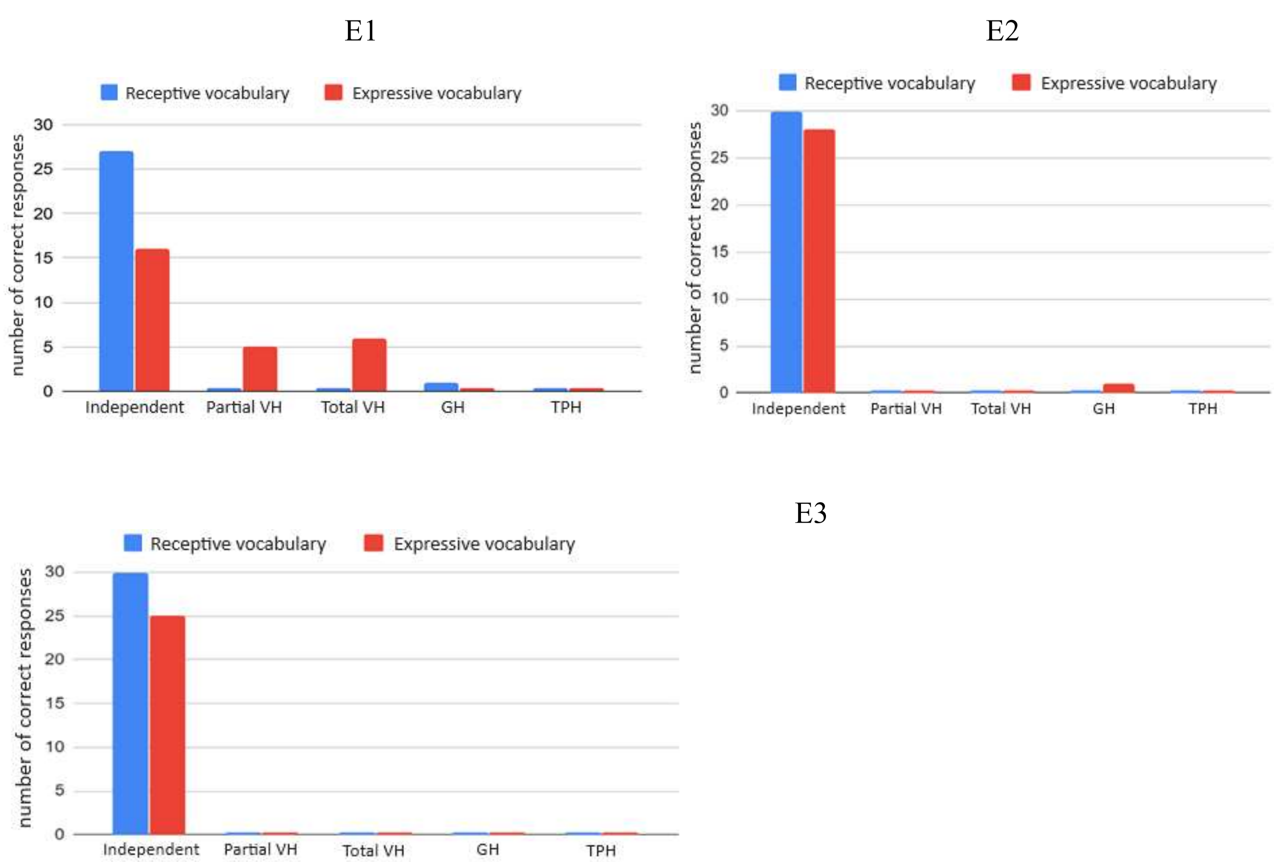

Figure 2. Performance of the three students in tasks characterizing receptive and expressive vocabulary. Caption: Independent = Correct response without hint; Partial VH = Partial verbal hint; Total VH = Total verbal hint; GH = Gestural hint; TPH = Total physical hint. Total possible correct answers for each type of vocabulary = 30.

Table 3. Gaze presence by delay time for E1, E2 and E3

Stimuli	Delay time	Number of evaluated trials		Average time spent in saccadic movements (%)		Standard deviation	
Gaze presence by delay time for E1							
Nonsocial	20 min	6		52,4		45,0	
Social	0 s	20	Total = 60	56,5	41,08	45,9	42,18
	4 s	20		52,3		41,0	
	8 s	20		14,5		24,6	
	20 min	6	0	0			
Gaze presence by delay time for E2							
Nonsocial	0 s	20	Total = 60	8,44	20,5	24,5	34,8
	4 s	20		52,4		38,9	
	8 s	20		0,565		2,53	
Social	0 s	20	Total = 60	2,97	2,83	10,3	10,4
	4 s	20		2,38		6,02	
	8 s	20		3,14		14,1	
Gaze presence by delay time for E3							
Nonsocial	0 s	20	Total = 20	74,5	63,01	20,1	29,14
	4 s	20		52,4		30,4	
	8 s	20		62,2		32,4	
	20 min	6		0		0	
Social	0 s	20	Total = 60	3,70	1,23	8,05	4,89
	4 s	20		0		0	
	8 s	20		0		0	
	20 min	6		0		0	

*Due to a procedural error, the data for the non-social stimuli at delays of 0, 4, and 8 seconds were not analyzed for E1. The data for the 20-minute delays were not analyzed for E2.

groups, demonstrating that the difference between the means observed for the two groups probably cannot be attributed solely to chance. And lastly, for E3, the t-test showed that there was a significant difference ($t = 16.194$, $p < 0.00001$) in gaze presence averages between social and non-social stimuli (Table 3), which could possibly be attributed to low detection of the gaze signal by the sensor during assessment with social stimuli.

Figure 3 shows the percentages of saccadic movement time in each attempt with each delay time, as well as the number of accumulated correct responses for each programmed delay scheme (n = 0s, 4s, 8s, and 20 min) for non-social and social stimuli. An ANOVA test was applied to analyze if there was a significant difference between the means of gaze detection time on the screen for each delay time. The ANOVA indicated a significant difference between the means of time spent in saccadic movements at each delay ($F(2, 57) = 7.313$, $p = 0.00149$) for E1, meaning that the difference between the means probably cannot be attributed to chance, with a 95% confidence level. The Tukey post-hoc test indicated that this was due to the difference between 8s delay and 0s delay ($p = 0.00283$) and 4s delay ($p = 0.00769$).

For non-social stimuli, to which item a of the figure refers, ANOVA indicated a significant difference ($F(2, 57) = 22.18$, $p < 0.00001$) in the percentage of gaze presence for the different delay times for E2 (Figure 3). The Tukey post-hoc test indicated that this difference occurred between the 4s interval and the other intervals (both with $p < 0.00001$). For social stimuli, ANOVA did not indicate a significant difference ($F(2, 57) = 0.029$, $p = 0.972$) in this percentage. Possibly, these results can be attributed to the non-detection of the gaze signal by the sensor in some parts of the task.

E3's performance (Figure 3) for non-social stimuli did not show significant difference ($F(2, 57) = 3.085$, $p = 0.0534$) in the percentage of gaze presence for the different delay times. For social stimuli, ANOVA indicated a significant difference ($F(2, 57) = 4.216$, $p = 0.0196$), which could possibly be attributed to non-detection of the gaze signal by the sensor during pairings with delays of 4 and 8 seconds, as the difference occurred exactly between these delays and 0s delay (according to the Tukey post-hoc test, with $p = 0.038$ in both situations).

Regarding the number of hits and misses (Figure 3), for non-social stimuli, E2 responded correctly to 2 attempts for a 0s delay; for 4s he responded correctly to 4 attempts; for 8s, he obtained 3 correct responses. For social stimuli, with a 0s delay, E2 responded correctly to 2 attempts; for 4s he responded correctly to 5 attempts; for 8s, he obtained

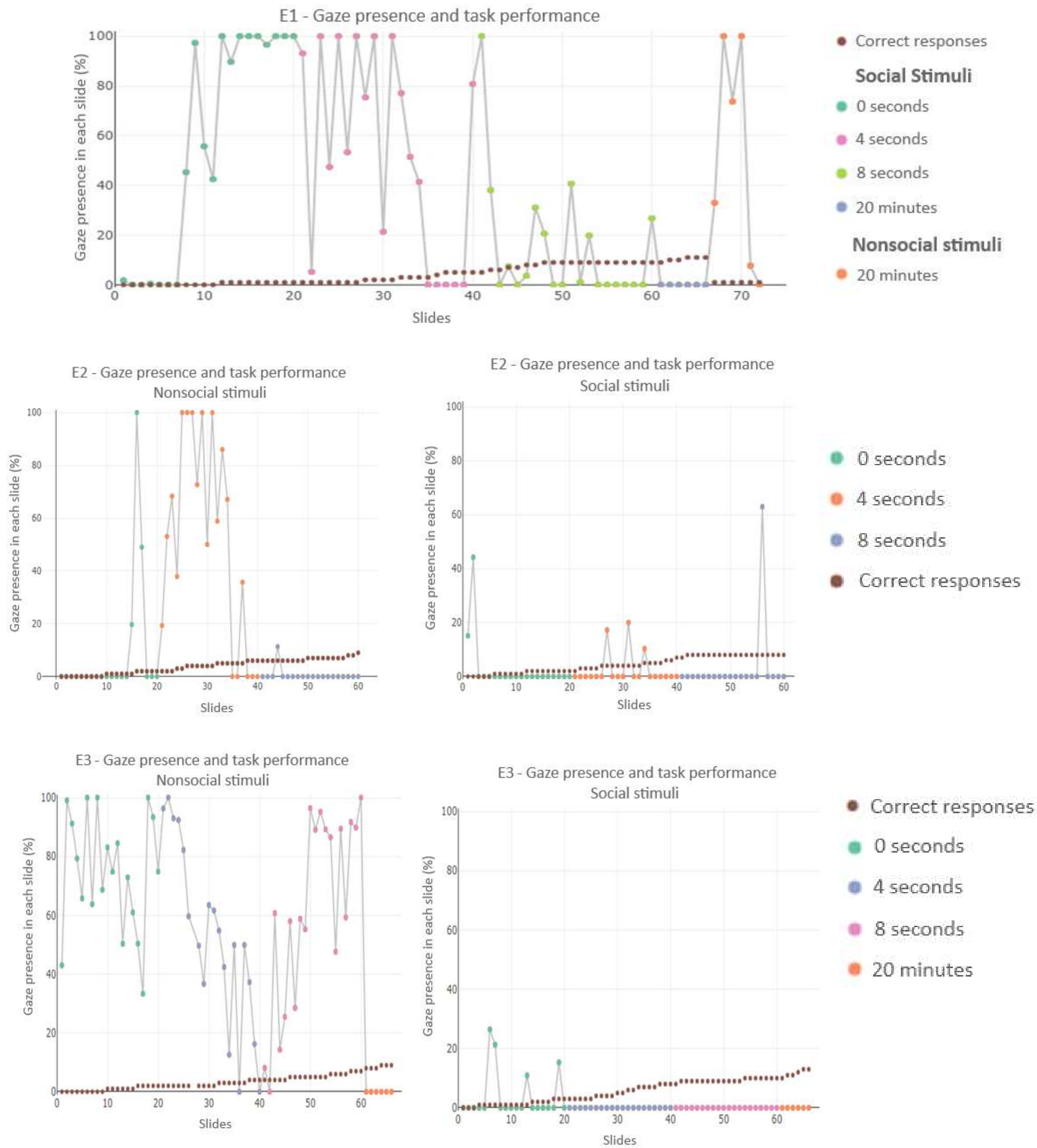

Figure 3. Cumulative number of correct responses for E1, E2, and E3 and percentage of gaze presence time (in relation to total time spent in a slide) at each delay time (0s, 4s, 8s, and 20 min), for the set of non-social and social stimuli.

one correct response. The maximum number of hits was 10 for all delays.

E3 (Figure 3), regarding non-social stimuli with a 0s delay, responded correctly to 2 attempts; for 4s, 2 attempts; for 8s, he obtained 3 correct responses. In all three delays, the maximum number of hits was 10. For the 20-minute delay, E3 responded correctly in 2 out of 6 attempts. For social stimuli, E3 responded correctly to 3 attempts for a 10s delay; for 4s, 5 attempts; for 8s, he obtained 2 correct responses. Finally, for the 20-minute delay, E3 obtained correct responses in 3 out of 6 attempts.

Eye tracking was analyzed based on heat maps and gaze trajectory (Figure 4 for E1, E2, and E3, based on the number and duration of eye fixations). It was found that even though all three selected the incorrect image, their gazes fixed on the correct image.

Figure 4. Eye tracking for E1, E2 and E3

E1's presented gaze tracking (Figure 4) refers to the second (E1a and E1b) and seventh (E1c and E1d) attempt to retrieve the previously presented image, both with a 0s delay. Items E1a and E1c show the heat map of eye fixation, where red represents a longer gaze fixation time on the region in question, and blue represents a shorter fixation time. Items E1b and E1d show the gaze trajectory, and the numbers in E1d indicate the order in which the movement occurred. In the second attempt, presented in E1a and E1b, gaze presence was detected 0.4% of the time the images were presented, with no fixation registered. In the seventh attempt, presented in E1c and E1d, gaze presence was detected 100% of the time. In this case, the image on the right was correct for the current attempt, and there was a greater fixation on it, as shown in the heat map, even though the student clicked between the two images (counted as an error).

For E2, the gaze tracking for non-social stimuli shown in Figure 4 refers to the 22nd (E2a and E2b) and the eighth (E2c and E2d) attempt, with delays of 0 and 8 seconds, respectively. Items E2a and E2c show the heat map of eye fixation. Items E2b and E2d show the gaze trajectory, and the numbers indicate the order in which the movement occurred. In the 22nd attempt, presented in E2a and E2b, gaze presence was detected 11.3% of the time the images were presented, with only one fixation detected, on the correct image, even though E2 selected the incorrect image. In the eighth attempt, presented in E2c and E2d, gaze presence was detected 100% of the time. In this case, the image on the right was correct for the current attempt, but there was greater fixation on the image on the left. However, E2 touched the correct image. E2's gaze tracking for social stimuli refers to the 17th attempt (E2e and E2f) and the 28th (E2g and E2h), with delays of 4 and 8 seconds. In the 17th attempt, presented in E2e and E2f, gaze presence was detected 10.3% of the time the images were presented, with no fixation detected. In the 8th attempt (E2g and E2h), gaze presence was detected 62.9% of the time. In this case, the image on the right was correct for the current attempt, but there was greater fixation on the image on the left, selected by E2.

E3's gaze tracking (Figure 4) for non-social stimuli refers to the 22nd attempt (E3a and E3b) and the 30th (E3c and E3d), both with an 8s delay. In the 22nd attempt, shown in E3a and E3b, gaze presence was detected 14.3% of the time the images were presented, with no fixation detected. In the 30th attempt, shown in c and d, gaze presence was detected 100% of the time. In this case, the image on the left was correct for the current attempt with fixations on both images, but the student touched the image on the right, counting as an error. For social stimuli, tracking was performed on the 3rd (E3e and E3f) and 4th attempt (E3g and E3h), with a 0s delay. In the 4th attempt, gaze presence was not detected at any point during the time the images were presented. In the 3rd attempt, gaze presence was detected 26.4% of the time. In this case, the image on the left was correct for the current attempt, and gaze concentration was observed on these images, even though the student selected the other image.

DISCUSSION

This study aimed to analyze eye tracking during delayed matching-to-sample tasks, involving delays of 0, 4, and 8 seconds and 20 minutes, using social (faces) and non-social (objects) stimuli with students with ASD and/or ID. It is essential to emphasize that no comparison was made between participants, as the study was aimed at a case study analysis.

The type of task used in the study involved the manipulation of mental information necessary to assess working memory through verbal and visual-spatial input, expanding the number of tasks evaluated in previous literature for these purposes (Siquara *et al.*, 2014). Regarding the type of stimulus (non-social or social), the hypothesis was confirmed, as statistically significant differences were identified for E2 and E3 regarding gaze time to the screen for each stimulus group, identifying a greater tendency of gaze toward non-social stimuli (mean percentage of gaze presence time - E2=20.5% for non-social and 2.83% for social, and E3=63.01% for non-social and 1.23% for social stimuli). These data replicate findings from previous literature regarding gaze duration for social and non-social stimuli (Hauck *et al.*, 1998; Schwartzman *et al.*, 2015). For E1, although the procedure was applied, such analysis was not conducted since it was not possible to capture data with non-social stimuli. However, it is essential to discuss, especially regarding E3 data, that they may have presented peripheral gaze, where the equipment was unable to capture such ocular behavior, thus reducing the gaze tendency average for the social stimulus group.

Regarding gaze time to the screen as a function of delay time, statistically significant differences were particular to each student, as no explicit difference was identified for all of them. For E1, in the social stimulus group, a difference was identified in the 8-second delay compared to the 0-second and 4-second delays. This implies that the longer the delay (in this case 8 seconds), the shorter the gaze duration to the screen. For E2, in the non-social stimulus group, a statistically significant difference was identified for the 4-second delay. For E3, in the social stimulus group, a difference was detected for the 0-second delay. Overall, at all delays (Figure 3), a considerable variance was observed in the percentage of time with gaze presence on the screen, corroborating previous literature showing that ASD conditions present differences in gaze fixation patterns compared to students without ASD (Mercadante *et al.*, 2006; Orsati *et al.*, 2009; Schwartzman *et al.*, 2015).

The number of correct responses varied as a function of each delay as well (Figure 3). With 0 seconds delay, E1 correctly responded to one trial; for 4 seconds, correctly responded to 4 trials; and for 8 seconds, 4 correct responses. For all delays, the maximum total correct responses were 10. This may indicate that the procedure taught E1 how to behave in each trial, suggesting the emergence of a learning set effect (Harlow, 1949; Saunders & Spradlin, 1993), where the student learned from the procedure which response to present in each contingency, for each trial, demonstrating a gradual repertoire acquisition. Still, regarding the relationship between the number of correct responses and delay time, no explicit trend was identified for all of them, and the initial hypothesis of relating task accuracy to longer gaze time to the screen was not confirmed, as none of them explicitly and consistently showed such a relationship throughout the procedure. The data did not replicate the effect of response and error for stimulus type, as in Hauck *et al.*'s (1998) study, and for delay time, as identified in Teixeira (2019).

In Hauck *et al.*'s (1998) study, students with ASD showed a higher number of correct responses in the non-social stimulus group compared to the social stimulus group. In the present study, the students showed a similar pattern regardless of the stimulus type. In Teixeira's (2019) study, the number of correct responses decreased as the delay increased, especially for students with ID/Down syndrome. In this study, no change in the number of correct responses was identified as a function of delay type.

E1 correctly responded to 9 stimuli at delays of 0, 4, and 8 seconds for social stimuli; E2 correctly responded to 9 non-social stimuli and 8 social stimuli at the same delays, and E3 presented 8 correct responses for non-social stimuli and 10 for social stimuli at such delays. E3 had the highest number of correct responses during delay tasks, differing from the data identified in McClain *et al.*'s (2022) study regarding the higher performance for the ASD group compared to the ID group.

Previous literature had already identified evidence of alterations in the pattern of eye movement of participants with ASD compared to those without the disorder, using tasks with delays of, for example, 20 minutes (Hauck *et al.*, 1998); however, questions related to ID and the type of stimulus to be used, such as social and non-social stimuli, were not compared. Thus, the study's contribution refers to the combination of two variables (delay time programming and type of stimulus to be used) in matching tasks with the auditory-visual arbitrary model, while evaluating both accuracy in task responses and implicit gaze measures (by computing the time the sensor detected gaze for each trial presented, the gaze heat map, and fixation position) with students with ASD and/or ID, in a natural data collection context. Evaluating the feasibility and conditions of data collection is essential for the improvement of future environments that may involve longitudinal studies of behavioral intervention.

The equipment used for analyzing eye movement was a portable model widely available on the market and of reduced cost compared to other eye tracker models used in research. However, one of the main limitations identified during this study was the difficulty in calibrating the equipment, especially with E1, as this required the student to maintain direct visual contact with a specific point on the screen. The student only proceeded to evaluation with delays when completing the calibration phase. It is recommended that future studies further investigate the role of calibration in capturing tracking data, in order to relate calibration performance to the time required for it and its possible interference in data capture.

Another limitation identified referred to the use of the touch-sensitive screen for conducting the experiment and the act of touching the notebook screen having created a physical barrier between the tracking equipment and the students' eyes, hindering data capture. However, the fact that the chosen equipment was portable enabled the tasks to be performed in the students' everyday environments, making data collection more naturalistic and showing applied possibilities for collecting this type of data.

The use of a portable device was challenging because, despite the application of all attempts provided in the procedure (Table 3), the sensor did not capture the gaze for all trials applied, and in this case, data analysis was performed with a different number of trials for each student. For example, the equipment did not capture any data out of 30 pairs of non-social stimuli for E1, nor out of the 12 stimuli of the 20-minute delay for E2. Thus, data analysis was conducted based on this type of variability. The use of non-portable equipment may minimize this type of variable due to higher sensitivity.

Recommendations for future studies include expanding the sample size and reevaluating the characterization of expressive and receptive vocabulary with a post-test intentions to verify if exposure to the task was sufficient to increase the correspondence between spoken word and corresponding image, as well as the analysis of the number of fixations for correct and incorrect stimuli in each trial. The use of echoics during the appearance of the stimulus for the first time (first screen of Figure 2, requiring student observation response) is suggested, given its function as a facilitator in the emergence of verbal operants (Costa & Souza, 2020).

Thus, the study highlights possibilities for collecting implicit measures in a natural situation, using a portable sensor, involving three students with ASD and/or ID. Overall, the manipulation of non-social and social stimuli and its relation to gaze presence were replicated, while the results of the number of correct responses did not show differences in performance between stimulus groups, and gaze presence did not interfere with trial accuracy and error. The findings indicate contributions to intervention planning with this population, especially related to the discrimination of social stimuli and the need to teach more efficient visual scanning, allowing for a more accurate discrimination of faces and facilitating social interactions.

REFERENCES

APA. (2013). American Psychiatric Association. *Manual diagnóstico e estatístico de transtornos mentais: DSM-5*. 5.ed. Porto Alegre: Artmed. http://dx.doi.org/10.5007/1807-1384.2014v11n2p96

Altenmüller-Lewis, U. (2017). Designing Schools for Students on the Spectrum. *The Design Journal*, *20*(sup1), S2215-S2229. https://doi.org/10.1080/14606925.2017.1352738

Bortoloti, R., & de Rose, J.C. (2009). Assessment of the Relatedness of Equivalent Stimuli Through a Semantic Differential. *Psychol Rec 59*, 563–590. https://doi.org/10.1007/BF03395682

Bortoloti, R., & de Rose, J.C. (2012). Equivalent Stimuli are more Strongly Related after Training With Delayed Matching Than after Simultaneous Matching: A Study Using the Implicit Relational Assessment Procedure (IRAP). *Psychol Rec, 62*, 41–54. https://doi.org/10.1007/BF03395785

Czermainski, F. R., Bosa, C. A., & Salles, J. F. de. (2013). Funções Executivas em Crianças e Adolescentes com Transtorno do Espectro do Autismo: Uma Revisão. *Psico*, *44*(4), 518–525.

Coderre, E. *et al.* (2019). Implicit measures of receptive vocabulary knowledge in individuals with level 3 autism. *Cognitive and Behavioral Neurology*, *32*(2), 95-119. https://doi.org/10.1097/WNN.0000000000000194

Costa, M. R. C., & Souza, C. B. A. (2020). Aquisição de intraverbais em crianças com autismo: efeitos do pareamento de estímulos e respostas ecoicas. *Psicologia USP*, *31*, e190061. https://doi.org/10.1590/0103-6564e190061

Dawson, G., Carver, L., Meltzoff, A. *et al.* (2002) Neural Correlates of Face and Object Recognition in Young Children with Autism Spectrum Disorder, Developmental Delay, and Typical Development. *Child Development*, *73*(3), 700-717. https://doi.org/10.1111/1467-8624.00433

Dias, N. M., & Seabra, A. G. (2013). Funções executivas: desenvolvimento e intervenção. *Temas sobre Desenvolvimento, 19*(107), 206-212.

Gaviria, O., & Fitzgerald, O. (2014). Análisis del funcionamento de la memoria operativa en niños con transtornos en el aprendizaje. *Acta Colombiana de Psicología*, *17*(2), 81-90. https://doi.org/10.14718/ACP.2014.17.2.9

Harlow, H. F. I. (1949) The formation of learning sets. *Psychological Review*, *56*, 51-65. https://doi.org/10.1037/h0062474

Hauck, M., Fein, D., Maltby, N., Waterhouse, L., & Feinstein, C. (1998). Memory for Faces in Children with Autism. *Child Neuropsychology*, *4*(3), 187-198. https://doi.org/10.1076/chin.4.3.187.3174

Huang, M. X., & Bulling, A. (2019). SacCalib: reducing calibration distortion for stationary eye trackers using saccadic eye movements. *COGAIN @ ETRA'1*9, 25–28. https://10.1145/3317956.3321553

McClain, M. B., Golson, M. E., & Murphy, L. E. (2022). Executive functioning skills in early childhood children with autism, intellectual disability, and co-occurring autism and intellectual disability, *Research in Developmental Disabilities, 122*, 104169. https://10.1016/j.ridd.2021.104169

Maenner, M. J. *et al.* (2016). Prevalence of autism spectrum disorder among children aged 8 years – autism and developmental disabilities monitoring network, *11 sites*, *69*(SS-4), 1-12. http://dx.doi.org/10.15585/mmwr.ss6904a1

Manzanero, A. L. *et al.* (2011) Reconocimiento de Caras y Discapacidad Intelectual. *Portada de Anuario de Psicología Jurídica*, *21*, 41-48. https://10.5093/jr2011v21a4

Maranhão, S. S. A., & Pires, I. A. H. (2017). Funções executivas e habilidades sociais no espectro autista: um estudo multicasos. *Cadernos de Pós-Graduação em Distúrbios do Desenvolvimento*, *17*(1), 100-113. https://dx.doi.org/10.5935/cadernosdisturbios.v17n1p100-113

Mercadante, M. T., Macedo, E. C., Baptista, P. M., Paula, C S., & Schwartzman, J. S. (2006). Saccadic movements using eye-tracking technology in individuals with autism spectrum disorders: pilot study. *Arquivos de Neuro-Psiquiatria*, *64*(3a), 559-562. https://doi.org/10.1590/S0004-282X2006000400003

Muñoz, P. O. L. (2018). *Rastreamento de olhar e reconhecimento de emoções em crianças com transtorno do espectro autístico*. [Tese de Doutorado, Universidade de São Paulo].

Orsati, F. T. *et al.* (2009). Percepção de faces em crianças e adolescentes com Transtorno Invasivo do Desenvolvimento. *Paideia*, *19*(44), 349-356. https://doi.org/10.1590/S0103-863X2009000300009

Paula, C. S. *et al.* (2011). Prevalence of pervasive developmental disorder in Brazil: a pilot study. *Journal of Autism and Developmental Disorders*, *41*, 1738-1742. https://doi.org/10.1007/s10803-011-1200-6

Roberts, L. V., & Richmond, J. L. (2014). Preschoolers with Down syndrome do not yet show the learning and memory impairments seen in adults with Down syndrome. *Developmental Science*, *18*, 404-419. https://doi.org/10.1111/desc.12225

Saunders, K. J., & Spradlin, J. E. (1993). Conditional discrimination in mentally retarded subjects: programming acquisition and learning set. *Journal of the Experimental Analysis of Behavior*, *60*(3), 571-585. https://doi.org/10.1901/jeab.1993.60-571

Schwartzman, J. S., Velloso, R. L., D'Antino, M. E. F., & Santos, S. (2015). The eye-tracking of social stimuli in patients with Rett syndrome and autism spectrum disorders: a pilot study. *Arquivos de Neuro-Psiquiatria*, *73*(5), 402-407. https://doi.org/10.1590/0004-282X20150033

Siquara, G. M. *et al.* (2014). Tarefas que avaliam memória operacional na infância e adolescência: uma revisão sistemática da literatura. *Estudos de Psicologia, 19*(4), 258-267.

Teixeira, I. O. (2019). *Efeitos do atraso em tarefas de discriminação condicional em participantes com e sem deficiência intelectual*. [Dissertação de Mestrado, Universidade Federal de São Carlos].

Zangrando, K. L. M. N. (2018) *Construção de um Paradigma de Rastreio Visual no Reconhecimento de Emoções em Crianças Autistas*. [Dissertação de Mestrado, Pontifícia Universidade Católica do Rio de Janeiro].

Zaqueu, L. C. C., Teixeira, M. C. T. V., Alckmin-Carvalho, F., & Paula, C. S. (2015). Associações entre Sinais Precoces de Autismo, Atenção Compartilhada e Atrasos no Desenvolvimento Infantil. *Psicologia: Teoria e Pesquisa*, *31*(3), 293-302. https://doi.org/10.1590/0102-37722015032243293302

Conflict of interest

The authors have no conflicts of interest to declare.

Data availability statement

The data supporting the findings of this study can be requested from the corresponding author upon reasonable request.

Funding information

This study was partially supported by the National Council for Scientific and Technological Development (CNPq). The work is the result of research and partnerships developed within the scope of the National Institute of Science and Technology on Behavior, Cognition, and Teaching (INCTECCE), funded by the National Council for Scientific and Technological Development (CNPq) (Grant No. 573972/2008-7) and by FAPESP (Grant No. 2008/57705-8). We thank Professor Deisy das Graças de Souza, president of the Institute, for her encouragement and support in conducting and publishing this study.

Corresponding author

Heloísa D'Angelo

Email: heloisadangelo@gmail.com

Submitted on

23/05/2022

Accepted on

13/04/2023