MACKENZIE PRESBYTERIAN UNIVERSITY

PAULO GUIRRO LAURENCE

INDIVIDUAL DIFFERENCES IN COGNITIVE ABILITIES AFFECTS STRATEGIES REVEALED BY EYE MOVEMENTS IN TASKS WITH ANSWER CHOICES

São Paulo

2021

PAULO GUIRRO LAURENCE

Individual Differences in Cognitive Abilities affects Strategies revealed by Eye Movements in Tasks with Answer Choices

Thesis submitted to obtain a doctoral degree in Developmental Disorders for the Postgraduate Program at Mackenzie Presbyterian University

Advisor: Prof. Dr. Elizeu Coutinho de Macedo

São Paulo

2021

Elaborado pelo Sistema de Geração Automática de Ficha Catalográfica da Mackenzie
com os dados fornecidos pelo(a) autor(a)

L379i Laurence, Paulo Guirro

INDIVIDUAL DIFFERENCES IN COGNITIVE ABILITIES AFFECTS STRATEGIES REVEALED BY EYE MOVEMENTS IN TASKS WITH ANSWER CHOICES: [recurso eletrônico] / Paulo Guirro Laurence.

3675 KB ; il.

Tese (Doutorado em Disturb do Desenvolvimento) - Universidade Presbiteriana Mackenzie, São Paulo, 2021.

Orientador: Prof. Dr. Elizeu Coutinho de Macedo

Referências Bibliográficas: f.98 -112

1. Working Memory. 2. eye Movements. 3. Eye-tracking. 4. Intelligence. 5. Reading. I. Macedo, Elizeu Coutinho de, *orientador*.II. Título.

Bibliotecário Responsável: Andrea Alves de Andrade - CRB 8/9204

Folha de Identificação da Agência de Financiamento

Autor: Paulo Guirro Laurence

Programa de Pós-Graduação *Stricto Sensu* em Distúrbios do Desenvolvimento

Título do Trabalho: Individual Differences in Cognitive Abilities affects Strategies revealed by Eye Movements in Tasks With Answer Choices

O presente trabalho foi realizado com o apoio de [1]:

☒ CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

☐ CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico

☒ FAPESP - Fundação de Amparo à Pesquisa do Estado de São Paulo

☒ Instituto Presbiteriano Mackenzie/Isenção integral de Mensalidades e Taxas

☐ MACKPESQUISA - Fundo Mackenzie de Pesquisa

☐ Empresa/Indústria:

☐ Outro:

[1] **Observação**: caso tenha usufruído mais de um apoio ou benefício, selecione-os.

PAULO GUIRRO LAURENCE

INDIVIDUAL DIFFERENCES IN COGNITIVE ABILITIES AFFECTS STRATEGIES REVEALED BY EYE MOVEMENTS IN TASKS WITH ANSWER CHOICES

Tese apresentada ao Programa de Pós-Graduação em Distúrbios do Desenvolvimento da Universidade Presbiteriana Mackenzie, como requisito parcial à obtenção de título de Doutor em Distúrbios do Desenvolvimento.

Aprovado em 04 de Novembro de 2021.

BANCA EXAMINADORA

Prof. Dr. Elizeu Coutinho de Macedo
Universidade Presbiteriana Mackenzie

Prof. Dr. Paulo Sérgio Boggio
Universidade Presbiteriana Mackenzie

Leandro de Castro

Prof. Dr. Leandro Nunes de Castro Silva
Universidade Presbiteriana Mackenzie

Prof. Dr. Alexandre Luiz de Oliveira Serpa
Meta Cognitiv

Prof.ª Dr.ª Sabine Pompeia
Universidade Federal de São Paulo

To my mother, Rose Guirro.

ACKNOWLEDGMENTS

Finishing a Ph.D. is a big achievement. Although there is only one name in the front cover of this thesis, it would be naïve to think that it was made by only one person. In this space, I would like to thank all the people who helped me achieve this goal. Since not all the people who will be mentioned here speak English, I wrote most of this section in Portuguese.

Primeiramente, eu gostaria de agradecer a minha mãe, Rose Guirro. Você é o meu porto seguro. Também gostaria de agradecer ao meu pai, Michel Laurence, por ter me ensinado muita coisa em tão pouco tempo, e entre essas coisas, ter me ensinado a ter paixão pelo ofício. Também gostaria de agradecer ao meu irmão, Felipe Laurence, e a minha tia, Jacqueline Laurence.

Esse trabalho nunca teria sido realizado se não fosse pelo meu orientador, Prof. Dr. Elizeu Coutinho de Macedo. Não tenho palavras para descrever o pesquisador que o Professor Elizeu é, e muito menos palavras para descrever a pessoa que ele é. O apoio em todas as minhas empreitadas e o incentivo durante as vezes que me encantei com algo na pesquisa foi fundamental para minha formação profissional e pessoal. O professor Elizeu me ajudou a superar meus limites. Na teoria da resiliência há um conceito chamado *turning point*, que é um momento em que a vida de uma pessoa sai de uma situação negativa para começar uma jornada positiva. Hoje, olhando para trás, tenho certeza de que nós termos nos conhecidos foi um *turning point* para mim.

Aqui também gostaria agradecer ao Prof. Dr. Paulo Sérgio Boggio por participar da banca de qualificação e de defesa desse trabalho, e por seus comentários que ajudaram a desenvolver o presente projeto. Também gostaria de agradecer pela convivência e pelos ensinamentos diários no laboratório.

Quero agradecer ao Dr. Alexandre Luiz de Oliveira Serpa por prontamente aceitar participar da banca da defesa dessa tesa e por todos os seus comentários. Além disso, agradeço pela convivência no laboratório que me ajudaram bastante na minha formação.

Agradeço a Profa. Dra. Sabine Pompeia e ao Prof. Dr. Leandro de Castro Nunes pela participação na minha banca de defesa do doutorado. Fico muito honrado de ter a presença de profissionais tão qualificados na banca e os *feedbacks* dados foram extremamente valiosos. Também agradeço a Profa. Dra. Katerina Lukasova por participar da banca de qualificação também, seus comentários foram muito importantes para o avanço do trabalho e nos deram uma nova perspectiva para o trabalho.

Esse trabalho também teve a ajuda direta de diversas pessoas, seja com a língua, com a coleta de dados ou com a elaboração. Nesse sentido, gostaria de agradecer à Mayara De Sousa Crippa Vaz de Oliveira Barros e Natalia Penteado Bertolino pelo auxílio na coleta de dados. Também sou muito grato à Stella Andrade Bassetto pela extensa ajuda com o inglês desse trabalho. Por fim, gostaria de agradecer à Tatiana Abraão Jana por disponibilizar parte dos seus dados.

Quero agradecer os meus colegas do Laboratório de Neurociência Social e Cognitiva da Universidade Presbiteriana Mackenzie que fazem o dia a dia na pesquisa ser muito mais cativante. Entre eles, gostaria de agradecer especialmente Gabriel Gaudêncio Rego e Lucas Murrins Marques.

I would also like to thank Prof. Silvia A. Bunge, Ph.D., for including me in the Bunge lab meetings and for helping me to develop the study presented in Chapter 4. I cannot express how valuable all of our meetings have been. I would also like to thank the Bunge lab members for being so welcoming. In particular, I would like to thank Elena Leib, Monica Ellwood-Lowe, and Willa Voorhies for their feedback on my project and data analysis.

Também gostaria de agradecer aos meus amigos Lucas Vasconcellos Coutinho, André Benedetti, André Pereira Duarte, Lucila Maria Cabral, Roberto de Stefani Neto, Victor Alleva Belleti e Artur Locatelli Ott, A amizade de vocês é um bem de valor inestimável para mim.

É necessário agradecer a Universidade Presbiteriana Mackenzie, seus professores e funcionários, uma vez que eles ofereceram uma ótima estrutura para os alunos, bem como funcionário extremamente qualificados.

Por fim, gostaria de agradecer a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) por financiar o meu doutorado.

"A brilliant strategy is, certainly, a matter of intelligence, but intelligence without audaciousness is not enough".

(Garry Kasparov)

Funding:

Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP (process number 2018/09654-7)

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES

Instituto Presbiteriano Mackenzie

ABSTRACT

When performing complex activities, it is common to use different cognitive strategies to solve the problem. These strategies can be accessed by various methods, such as self-report, observation of steps during the task, or eye movements. The application of the method depends on the type of activity. Tasks that require one step at a time from the participant make it possible to analyze the task step by step. Tasks, where there are no clear steps, require other methods. An example is tasks with answer banks where the test-taker must choose an option, such as matrix reasoning or gap-fill tests. In this type of task, visual inspection of the test is critical for a correct answer, in addition to revealing which cognitive strategy was used. Furthermore, these strategies are mediated by diverse cognitive abilities. With that in mind, the objective of this thesis was two-fold: we aimed to analyze which cognitive strategies, measured by eye gaze patterns, are used in different tasks with answer banks; secondly, we aimed to relate the cognitive strategies with several cognitive measures in order to pinpoint which cognitive abilities are related to these strategies. This thesis has three studies: a theoretical review of the cognitive strategies applied in matrix reasoning tests, and two experimental studies. The first study had two samples (N = 62 and N = 73) and aimed to explore the relationship between cognitive abilities and eye-tracking measures related to strategy use in matrix reasoning tasks Results indicated that self-reported executive functions, working memory, and planning are directly related to the strategies applied in the matrix reasoning test. Fluid reasoning also showed a relationship with the applied strategy, but less than the other cognitive abilities. The second study had a sample of 51 participants and the objective was also two-fold: first, we aimed to identify the cognitive visual strategies in the cloze test using a non-supervised algorithm that accounts for the scanpath instead of the summarized events; second, we aimed to analyze the relationship of these strategies with the working memory and the performance in the cloze test. It was possible to category the eye gaze in two strategies: a global and a local strategy. There was a direct relationship between working memory and which strategy was used, with the global strategy being related to individuals with higher working memory. Furthermore, the cognitive-visual strategies were also related to performance within the Cloze test: the global strategy was related to better performance. With the results of the three studies, we were able to understand that working memory is a fundamental cognitive component that mediated which cognitive strategies were used by test-takers. In addition, patterns of high alternance between different areas of interest in tests are usually related to poorer performance, which may indicate that when participants are able to perform what is necessary to solve the

test item, they do it directly and concisely, but when there is some difficulty, they make multiple comparisons between different areas trying to find a missing clue. These findings may guide future research in understanding cognitive strategies on different tests, as well as interventions for groups that are underperforming.
Keywords: Working Memory, Eye movements, Eye-tracking, Cognitive Function, Intelligence, Reading.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AOI – Area of Interest

AUC – Area Under the Curve

BF – Bayes Factor

BRIEF-A – Behavior Rating Inventory of Executive Function for Adults

EF – Executive Functions

Gf – Fluid reasoning

IES – Inverted Efficience Score

LASSO - Least Absolute Shrinkage and Selection Operator

M – Mean

MAE – Mean absolute error

ms – Milliseconds

MTDI – Matrix time distribution index

RAPM – Raven Advanced Progressive Matrices

RMSE – Root-mean-square error

SD – Standard Deviation

WCST – Wisconsin Card Sorting test

WM – Working Memory

WMC – Working Memory Capacity

WMT-2 – *Wiener Matrizen Test 2*

TABLE OF CONTENTS

PRESENTATION

The project executed here was developed at the beginning of September 2020, due to the COVID-19 pandemic. In fact, this was the present student's third Ph.D. project since March 2020. The present student had to cancel the first project because it became unfeasible to collect data due to the pan demic outbreak, then developed a second project to collect data online, but the data resolution was not enough for the project's aim.

The first project involved data collection with electroencephalography and eye-tracking, but it was not possible to do it due to the university closing due to the COVID-19 pandemic, in addition to possible risks to participants and the researcher. The second project had the intention of making online collections with eye-tracking via webcam. However, the technology has not proven itself reliable enough for what the project needed. So, the project that produced this present work came about. The data, which will be presented later in Studies 2 and 3 were collected in the laboratory and were collected for other projects, that were carried out before the COVID-19 pandemic but were never published or even analyzed in an in-depth and systematic way.

This work brings much of the projects that the present student has engaged in recent years, and has a direct continuity with the work developed in his graduation and master's thesis. In addition, the student was connected with all of the collected data. This context enabled the formulation of hypotheses within the context of the studies that will be presented here. Therefore, the formulation of this project was natural since it became clear that the two previous projects could not be carried out. These ideas were already formulated in a non-formal way. With this given context, we can proceed to the introduction of the project.

CHAPTER 1: INTRODUCTION

Research on how the participant thought in order to solve a task is not new (e.g., Newell & Simon, 1972). However, these initial studies, from the 70s and 80s, were limited. Cognitive strategies were only possible to be analyzed in tasks that presented multiple steps, that is, tasks that demanded continuous steps from the participant so that the researcher could infer what the participant was thinking during the task (Groner & Groner, 1982, van der Maas et al., 2011). Thus, the study of cognitive strategies focused on very specific tasks that were not always ideal for understanding the underlying cognitive processes. But that has changed since the 90s.

Thanks to technological advances and the increased processing power of computers, it is now possible to use eye movement tracking techniques to see where participants are looking in a task (Cherubino et al., 2019, Płużyczka, 2018). With ever-increasing resolution in eye movement tracking technologies and the use of advanced data analysis techniques, it is currently possible to extract data from a participant's eye movements during a task and then process it to obtain numerical measurements that can be used to infer the cognitive-strategic pattern in a test (e.g., Ito et al., 2017, Kucharský et al., 2020, Mccray & Brunfaut, 2018, Vigneau et al., 2006). Until the 90s it was extremely difficult to obtain quantitative measures of cognitive-visual strategies, being more common qualitative works of classification of visual tracing patterns (e.g., Bethell-Fox et al., 1984; Snow, 1978, 1980).

Since it is possible to use eye-tracking methods in different tasks, the study of cognitive strategies is not reserved for tasks with multiple steps anymore. But the eye-tracking method also has some requirements for the type of task used. In this case, the task must have different Areas of Interest. In other words, the task must have regions that the participant has to do a sequential eye gaze in order to understand what is necessary of them to complete the task. This is particularly important for algorithms that classify and cluster scanpath over time (Kucharský et al., 2020). In this sense, tasks with answer banks where the test-taker must choose an option are great options for analyzing strategy through eye movements. Examples of this type of task are the matrix reasoning task and the reading comprehension Cloze-type task. In the matrix reasoning tasks, there are two main areas of interest: the matrix and the answer choices. In the Cloze test, there are also two main areas of interest: the text and the word bank. Both of these tasks have similarities: the test-taker has to gaze at the matrix/text and then do comparisons with the answers choices/word bank in order to answer the test. For this reason, these texts are very fruitful for analyzing different cognitive visual strategies.

Studies on cognitive strategies drawn from the analysis of eye movements are a necessary effort to understand how underlying cognitive processing in different tests, but there are still few studies that fully address these issues. In fact, most of them are recent. In matrix intelligence tests, the first study identified in the literature was only in 2006 (Vigneau et al., 2006), and it took another five years for a new study to systematically address this issue again (Hayes et al., 2011). In reading comprehension tasks from cloze-type tests, the first article published working directly with this issue in a systematic way was only released in 2018 (Mccray & Brunfaut, 2018), although some studies have previously hypothesized about the topic (Gao & Gu, 2008, Yamashida, 2003). This indicates the frontier nature of the topic.

Although scientific knowledge has been produced in such areas, this knowledge still seems to be initial and is relatively scarce and poorly connected. Thus, there is a need for studies that deepen the understanding of these phenomena. It is very important to understand what these strategies in different tasks have in common with each other and to know if they share the same underlying cognitive processes. In other words, it's important to understand what unites them as strategies.

Thus, the purpose of this thesis was, above all, to tell a story: a story about how we apply cognitive strategies, which are possible to be captured by pattern in the eye gaze, in different tasks that seek to achieve a clear objective. Further, we aimed to understand what relationships these strategies have with each other. And most importantly, it is also necessary to understand which underlying cognitive processes these strategies are based on. Putting this in a systematic way, our objective was two-fold: firstly, we aimed to analyze which cognitive strategies, measured by eye gaze patterns, are used in different tasks with answer banks; secondly, we aimed to relate the cognitive strategies with several cognitive measures in order to pinpoint which cognitive abilities are related to these strategies.

Different cognitive abilities can be related to these strategies. Some examples are working memory, planning, cognitive flexibility, and executive functions. All of these abilities are important when thinking about cognitive strategies: whether to store and memorize task patterns; to inhibit useless or stimuli with poor information during the test; or even more to give up an ineffective strategy to venture into new strategies.

In fact, some studies have deliberately shown the relationship between working memory and these strategies in specific tasks (e.g., Gonthier & Roulin, 2019, Gonthier & Thomassin, 2015, Jarosz et al., 2019) or have hypothesized about the subject (e.g., Mccray & Brunfaut, 2018). However, no study has gone further, trying to understand the relationships of these strategies with other cognitive abilities. Thus, this issue was, to some degree, addressed in this thesis.

CHAPTER 2: THEORETICAL FRAMEWORK

Cognitive abilities

Cognitive abilities are defined as mental broad abilities that are related to reasoning, planning, problem-solving, working memory, inhibitory control, cognitive flexibility, learning, comprehension knowledge, abstract thinking, among others (Gottfredson, 1997a, Ispas & Borman, 2015). These abilities present individual differences, meaning that each person will have different and unique levels for each ability (Ones et al., 2012).

Several models were created to comprehend the relations of cognitive abilities, but two of the most extensive models are the CHC model (Carroll, 1993, McGrew, 2009) and the Miyake and Friedman's model (Miyake & Friedman, 2012). Both models are based on factorial analysis of different cognitive abilities (Miyake & Friedman, 2012, Ones et al., 2012).

The first model was based on Carrol's Three Stratum theory, and it dissects the cognitive abilities in narrow and broad abilities, with the narrow abilities being the first stratum, the broad abilities being the second stratum, and the third stratum is defined by the general intelligence (Carroll, 1993, McGrew, 2009, Ones et al., 2012). This model is based on the Cattell-Horn Gf-Gc model, a model that separates the ability of reasoning (Gf), that does not need any previous knowledge, and the accumulation of knowledge (Gc; Carroll, 1993, Schneider & McGrew, 2012). Both Gf and Gc are inserted in the second stratum of the model, with other abilities, such as short-term memory, visual processing, auditory processing, long-term memory, reading and writing, among others (Schneider & McGrew, 2012). These abilities are second-order factors of different narrow abilities measured by psychological tests, the first stratum. The second stratum also relates itself to the third stratum. In this stratum, there is only general intelligence (McGrew, 2009, Schneider & McGrew, 2012).

The second model is an Executive Function (EF) model. One definition of EF is that they are abilities that help a person perform a task. For this, it involves aspects such as manipulating focus, solving problems, organizing, making decisions, correcting mistakes, planning, creating goals and following them, and, finally, avoiding external distractions (Diamond, 2013). Executive functions manifest themselves in goal-directed behaviors that require us to deal with multiple pieces of information at the same time. Because of this, they impact academic and professional performance and have several consequences in life (Bailey, 2007, Guatercole et al., 2008, Guatercole & Pickering, 2000).

Miyake and Friedman's model of EF proposes that there are three main specific components, namely: updating, the ability to monitor and manipulate different information in the working memory; shifting, the ability to switch between tasks or mental states; inhibition, the ability to deliberatively overrule an automatic or dominant response. The authors also suggest that there are other potential EFs, but that updating, shifting, and inhibition presents good insights into the individual differences among EFs. This model also states that the EFs have a common underlying ability, but each of the EF has its uniqueness and independence (Miyake & Friedman, 2012).

Although there are differences between each cognitive abilities model, some studies (e.g., Buczyłowska et al., 2020, van Aken et al., 2019) presented a strong relationship between the cognitive abilities that are common in the CHC model and the EFs. But it is also noted that some specific abilities, such as planning and inhibition, have their own variance, and it is not a good predictor of general intelligence (van Aken et al., 2019).

The cognitive abilities present real word impacts. For example, cognitive abilities are considered one of the biggest predictors for job performance (Schmidt & Hunter, 1998), and higher cognitive abilities are related to more complex work tasks (Gottfredson, 1997b).

All these cognitive abilities can be measured in different tasks. However, the way that each person finds a solution to these tasks is different. Therefore, it is necessary to understand the strategies that they apply in order to find a solution.

Strategies in cognitive tasks

Strategies employed during the performance of a cognitive task occur as ways to find a solution to the task. These strategies reflect which elements were understood by the participant and the way he thought while trying to solve the task (Steingroever et al., 2019). Thus, cognitive strategies can be defined as mental methods and schemas that individuals use to solve problems.

One of the lines of research involving strategy is what leads a person to adopt one strategy or another (Siegler, 1988). As a branch of individual differences, two variables were shown to be important for the choice of different strategies: intelligence (Bexkens et al., 2016) and inhibition capacity (Borst et al., 2012, Poirel et al., 2012).

The second line of research involving strategies tries to understand the change in strategies according to human development. Changes in strategy resulting from development are found in decision-making studies (Bereby-Meyer et al., 2004, Betsch & Lang, 2013, Huizenga et al., 2007, Kwak et al., 2015), reasoning (Jansen, Van der Maas, 2001, 2002, Siegler

1987; 2007, Van der Maas & Molenaar, 1992), mathematics (Ashcraft & Fierman, 1982, Bjorklund & Rosenblum, 2001, Cho et al., 2011, Torbeyns et al., 2009) and categorization (Rabi et al., 2015; Rajumakers et al., 2004).

Strategies are primarily evaluated by behavioral measures. In fact, with measures of time, success, and which item was chosen (or error analysis), it is possible to infer which strategy the person was doing during the test (Groner & Groner, 1982, Van der Maas et al., 2011). For tasks with multiple steps (e.g., Tower of Hanoi, Tower of London, Balance Scale Task, Iowa Gambling Task), in which several steps are needed for the participant to reach the goal, it is possible to analyze each participant's movement and infer his strategy.

Following this thought, Newell and Simon (1972) used the Tower of Hanoi test to study these strategies. Based on the participants' behavior, they were able to analyze sequences of behaviors in the test to the point of categorizing them into different groups. These groups, called heuristics by the authors, would be the possible strategies applied to solve the problem. Below, we will present some heuristics found by Newell and Simon (1972) and other authors.

Means-end analysis: is a heuristic in which the participant understands the difference between the current problem and the goal and then uses sub-goals to bring the current problem closer to the goal. This heuristic assumes that the participant knows the goal. While it can be effective for some simple tasks, especially at an early stage, for other tasks it can be quite damaging. By persisting in this heuristic for a long time when it is not ideal for solving the problem, the participant may present great inefficiency in the task (Sweller & Levine, 1982). An example of this is the maze task by Sweller and Levine (1982). In this task, the participant starts in the middle of a square maze and has to reach the lower-left corner. The move indicated by this heuristic is to make small subgoals until you reach the correct point. However, the maze is constructed in such a way that the participant must reach the right apex and then traverse the upper and left edge of the maze until reaching the goal point. In this way, the participant who insists on paths to the left and down will have an inefficient performance and will need several moves to reach the goal, while a participant who does not submit to such heuristics will reach the goal much more quickly.

Mountain climb: This heuristic would be the simplest of all. It is used by participants who don't exactly understand the structure of the problem. It is simpler than means-ends analysis and would be comparable to a climber who wants to climb a mountain and to do so simply tries to follow a straight line to the summit of that mountain. Because of this, it is possible that it also has efficiency problems.

Progress Monitoring: This is a heuristic presented by MacGregor and colleagues (2001). This heuristic is based on the feeling of moving forward in solving a problem. If progress is extremely slow, the participant will tend to try different techniques to solve the problem. Evidently, in tasks that create the illusion of progress, the participants who used this heuristic would end up being less efficient.

Thus, different behaviors during a test are important because they demonstrate different cognitive strategies. These strategies, depending on the task, can help the participant to be more or less efficient and perform better or worse (e.g., Sweller & Levine, 1982).

However, there are tasks where it is not possible to directly analyze the participant's movements (e.g., matrix intelligence test; A:B::C:? analogy tasks; other multiple-choice tasks). Thus, it is not possible to deduce the strategy applied by the participant through the observable behavior during the task. Therefore, an excellent option to be able to infer the participants' cognitive strategies during the execution of a task in which it is not possible to analyze the participants step-by-step is the eye-tracker equipment. This technique allows the participant's eye movements to be recorded so that it is possible to infer which areas the participant looked at. From this, it is possible to deduce complex strategies based on the places the participant was looking at (e.g., Bethell-Fox et al., 1984; Laurence et al., 2018; Vendetti et al., 2017; Vigneau et al., 2006). Thus, it is necessary to talk about eye-tracking before we go deeper into the strategies.

The eye-tracking

Eye-tracking records began in 1879. Ewald Hering (1879) and M. Lamare (1892) were the first to describe eye movements during reading. These researchers carried out experiments in which participants had to read a text, and the researcher analyzed the participants' movements with the naked eye since there were no more developed technologies. In these experiments, researchers noticed that the participants did not read continuously with their eyes, but instead paused briefly in words and then quickly moved to other words. These quick movements were called “saccades”, one of the most notorious terms when we talk about eye movements.

By chance of fate, many works cite Louis Émile Javal as the first to describe eye movements, but this was possibly due to an interpretation error (Wade & Tatler, 2008). Javal (1878; 1879) commented on the work of Hering and Lamare that had not yet been published, since these researchers worked in the same laboratory as he, and because of this there was a misinterpretation that he was conducting the work (Wade, 2010, Wade & Tatler, 2008).

It is interesting to note that Javal was nevertheless the first to use the word “saccades” (Wade et al., 2003). Javal was French, and the word “saccades” comes from the French, which is used to refer to the rapid movements that horses make when being trained, but its translation for the English term “jerk” was little adopted by researchers at the time. The use of the term “saccades” until today demonstrates that although the first works on eye movements were not authored by Javal, he still influenced all works that use the register of eye movements (Wade, 2010, Wade et al., 2003).

Possibly the first work in the English language to cite the results of Hering and Lamare, described by Javal (1879), was the article by Edmund Huey (1898). Later, Huey (1908) developed an eye movement tracking system. This was possibly the first eye-tracker, which was an intrusive mechanism. Participants needed to wear a contact lens with a small opening for the pupil while that lens was connected to a pointer that changed position according to their eye movements. This mechanism helped to understand which words participants stopped while they were reading (Huey, 1908; Wade, 2010).

The first non-invasive eye-tracker that allowed the recording of horizontal and vertical movements was introduced by Judd and colleagues (1905). This mechanism worked with a mechanical indicator relative to the participant's eye that reflected a light point on a photosensitive film, allowing to record frame by frame where the participant was looking. Although this method had the positive side of not being invasive, it had a problem: the participant's head needed to remain completely still while he performed the task (Płużyczka, 2018).

Along with Judd, another researcher, Guy Buswell, began to make advances in eye-tracking studies. In some of his eye-tracking work, Buswell (1922; 1937) analyzed how adults and children read. His work was pioneering and identified differences in oral or silent reading. It was also possible to notice that there were differences due to age and educational level. In addition, his work has also demonstrated that it is possible for adults to learn to improve their reading and that many adults do not read as well as is expected of them. These works were essential for the development of future research in the field of reading (e.g., Rayner, 1998).

Buswell was also interested in how people look at pictures (Buswell, 1935). In his study, Buswell collected data from 200 participants, who watched several figures, totaling more than 2000 recorded data, a formidable number of data collections even for the present time, and extremely incredible for the technology of the time. In this frontier and classic work, the author sought to systematically explore eye movements as the subject looked at complex images rather than simple shapes or texts. In the study, Buswell sought to identify the distribution of fixations in the figures, how the first fixations are different from the last fixations, the duration time of the first fixations compared to the last ones, how fixations changed with time, what was in common between the tracings of several participants in the same figure, and finally, how the instructions given to the participant could change such tracing.

With this work (Buswell, 1935), Buswell was able to identify several issues, such as that there is a relationship where the person looks with the fixation time or that people with artistic training have different patterns than people without artistic training. However, this author realized that eye movement patterns were insufficient to broadly describe our visual experience. In this way, he pointed out the importance of explaining the vision cognitively. This was a great advance for eye-tracking studies, as it paved the way for more complex research to take place (Wade, 2010).

Buswell's works were also important to understand the difference in tracings due to different instructions. In this sense, he first let the participants look at the image freely and then asked them to look for certain things within that figure (Buswell, 1935). Another researcher later revived this question using a more elegant method and produced one of the most classic, if not the most classic, studies on eye-tracking. In this case, Alfred Yarbus (1967) presented a figure seven times, each with a different instruction. The tracings of each figure were extremely different, depending on the instruction given (Figure 2.1). Thus, the author was able to demonstrate that higher cognitive elements were able to inhibit lower cognitive elements (Tatler et al., 2010; Wade, 2010).

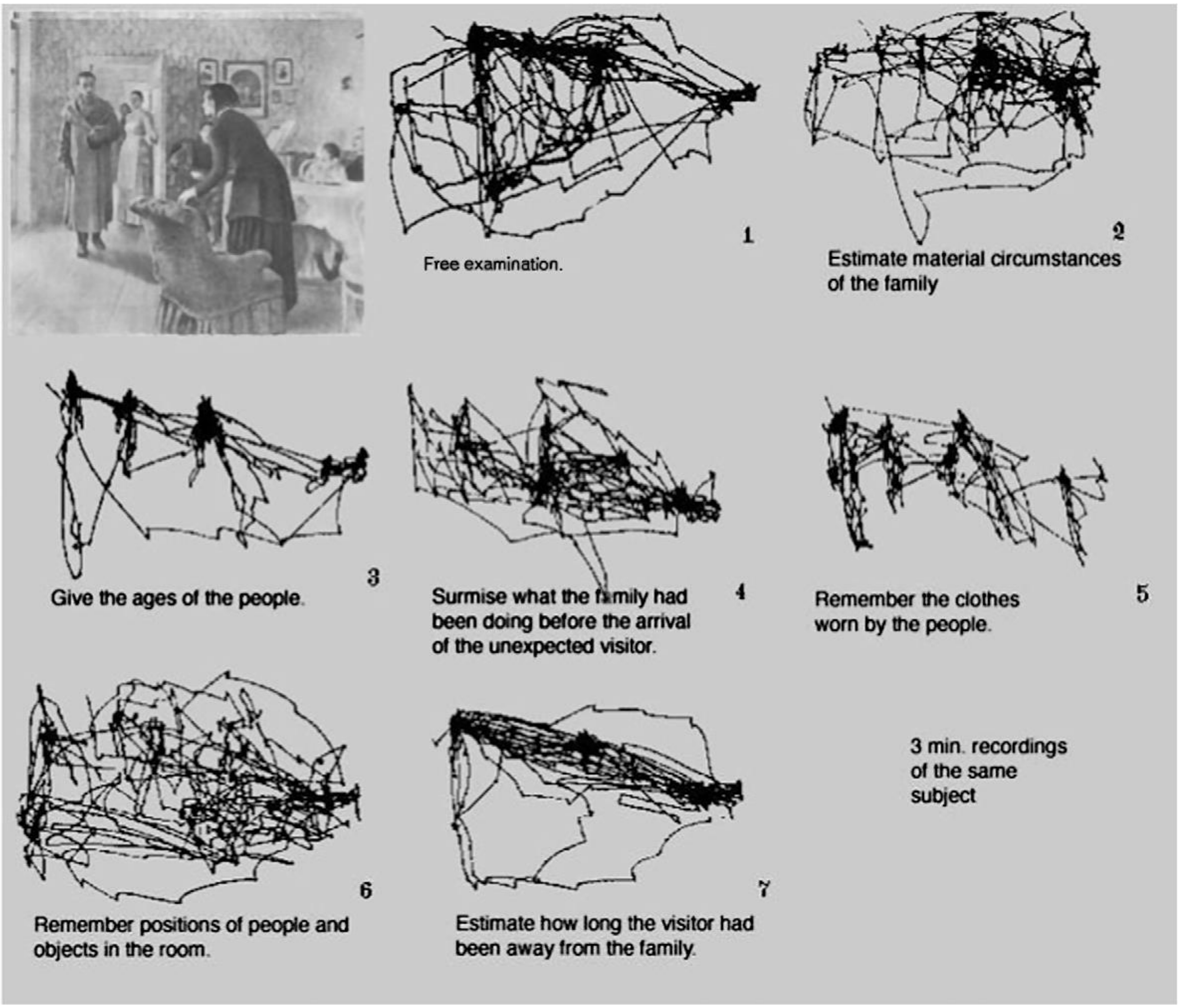

Figure 2.1 – The different eye movement tracings derived from distinct instructions found by Yarbus. Source: Yarbus (1967), adapted by Greene et al. (2012).

Another important point of Yarbus' (1967) works was that he realized one of the main foundations of studies with eye-tracking: during the perception of faces, the visual scanning takes place mainly in both eyes and mouth, and there is a triangulation between these three areas. These works are extremely relevant to cognitively understand the processing of faces and conditions related to this processing, such as autism (e.g., Papagiannopoulou et al., 2014).

Much of Yarbus's (1967) research was devoted to when the eyes are still. This phenomenon, called fixation, was investigated in depth by the researcher and is where most research on eye-tracking has been focused until today. One of the points that Yarbus (1967) tried to study was whether there was a relationship between involuntary eye movements, known today as microsaccades, and visual acuity. For this, Yarbus used a contact lens that was able to stabilize the eye through suction, and the results he obtained indicated that after 1 to 3 seconds there was a significant loss of visual acuity.

To study fixations and microsaccades in-depth, it was necessary to develop technologies for recording eye movements, since these movements are of high frequency and extremely short and fast. For this, equipment similar to those currently used was created. The first eye-tracking equipment similar to current ones was developed by the United States Air Force in the 1960s. It was made possible thanks to computer algorithms that allowed to identify the iris of the eyes and then process the point on the screen that the person was looking (Merchant et al., 1974; Płużyczka, 2018). At that time, some more advanced equipment had a recording rate of 30 to 70 Hz (Wade, 2010).

In the 1970s, new areas of eye-tracking research began to emerge, this time led by psychologists concerned with the foundations of cognitive psychology. This era coincided with the theories of language, which made the use of eye-tracking technologies extremely related to reading research until today (Płużyczka, 2018). One of the big names at this time was Keith Rayner, who published works that are now considered classics (e.g., Rayner, 1975, 1978, 1979; for a review, see Rayner, 1998 and Clifton et al., 2016).

Finally, a new era of research emerged in the 1990s and continues to this day. This era began with the development of eye movement recording technologies, at the same time that there was a huge boost in the data processing and operation capacity of computers. In this way, the eye movement recording equipment was more apt to process data that had noise due to some movements of the participant's head, allowing the participant to no longer need to have their head immobilized (Płużyczka, 2018). This great development in technology allowed eye-tracking to be popularized in several study topics such as psychology, neuroscience, marketing, economics, ergonomics, medicine, technology areas, among others (Cherubino et al., 2019; Płużyczka, 2018).

With the development of technology and the field of study itself, it was possible to better understand what eye movements really are. Thus, the saccades, baptized by Javal, are movements between regions of interest. The saccades occur between fixations, and the eyes move because of the peripheral vision that finds some element that attracts attention to the point where a new fixation occurs. In a scenario with several elements, the saccades occur due to a relationship between the objectives of the observer and the properties of the observed scenario. During the saccade, visual information is collected for the saccade to end. Furthermore, saccades occur through both eyes, in a conjunctive way: each eye presents the same direction and the same amplitude (Gilchrist, 2011).

Saccades have a very interesting feature: the movement starts from a quasi-steady state, peaks extremely quickly, and then returns to a new quasi-steady state in another space. Saccades can reach a maximum speed of 500º/sec and typically have a shorter duration of between 30 and 80 milliseconds (ms). The vast majority of saccades have variations of 4-20º. For larger saccades, head movement is used (Gilchrist, 2011; Holmqvist et al., 2011).

Where the saccade ends is an important factor. In laboratory situations, the saccade may end just before the target, making a new corrective saccade to adjust necessary. This phenomenon is known as hypometric saccade. In realistic environments, the observed phenomenon is different: usually saccades end exactly where they should end, even for long saccades. In geometric shapes, for example, the eye usually lands in the middle of the figure, showing great precision. In the case of moving targets, the saccades tend to have a corrective process, known as saccades adaptation, which helps the motor systems to adapt to a correct response (Gilchrist, 2011).

The saccade latency is the period until a new saccade starts. Some latencies are 100 ms while others exceed 1000 ms. In general, the latency distribution of saccades is Gaussian, with a long tail for longer latencies. This latency can be seen as a process: it boils down to a complex decision process that, when it crosses a threshold, initiates a new approach. Therefore, saccade latency is composed of the stimulus processing time, the accumulation of the decision process, and motor execution (Gilchrist, 2011).

This period of latency of the saccade largely involves moments when the eyes are still, a situation much studied by Yarbus (1967). Today, we know that the eye is never fully still. As Yaburs (1967) demonstrated, if the eyes are fixed, visual acuity is lost within 1 to 3 seconds. However, from a theoretical and historical point of view, this phenomenon is still called fixation, while some more recent terminologies use the term fixating eye movements (Martinez-Conde & Macknik, 2011, Martinez-Conde et al, 2004). In the real world, most of these fixations have durations of 200 to 300 ms (Holmqvist et al., 2011). In fact, our eyes spend 80 to 90% of our waking hours fixating on something (Duchowski, 2007, Martinez-Conde & Macknik, 2011). During this fixation period, the eyes make small movements that shift focus on the retina. This is because the nervous system is made to perceive changes, and with the oscillations of focus of the visual sensory system, it is possible to constantly generate small changes to the nervous system. This prevents static objects from disappearing due to the lack of changes (Martinez-Conde & Macknik, 2011).

Fixing eye movements, or fixations, are the least understood phenomena of all types of eye movements. They are made up of three types of movements: tremor; drifts; and microsaccades (Holmqvist et al., 2011; Martinez-Conde & Macknik, 2011; Martinez-Conde et al., 2004).

Simply put, tremors, or physiological nystagmus, are very small oscillatory movements (< 1'), high frequency (> 90 Hz), and not so fast (20'/sec maximum). Recording this movement is difficult because its frequency is similar to the noise found in eye movement recording equipment. There is no consensus on the role of these movements, but one of the possibilities is that it is imprecise muscle control. Drifts are slow movements (200 to 1000 ms), not so fast (6 to 25'/s), of deviation of the eye from the focus it was on. They have a range from 1 to 60'. This movement can occur for a compensatory issue, in order to avoid sensory adaptation (Holmqvist et al., 2011, Martinez-Conde et al., 2004).

Finally, microsaccades are short movements, with amplitudes from 10 to 40', fast (10 to 30 ms), relatively fast (15-50º/s), and involuntary. Because of these measurements, microsaccades are the fastest and widest fixating eye movements. Like saccades, they are also binocular, with very similar amplitudes and directions between the two eyes. They occur as a correction of drift, bringing the focus of the eyes back to its original position (Holmqvist et al., 2011, Martinez-Conde et al., 2004). Currently, there is evidence that microsaccades are related to perceptual, attentional, and cognitive issues. An example of this is that the production of microsaccades is possibly related to working memory (Valsecchi et al., 2007, Valsecchi & Turatto, 2009) and that the absolute frequency of microsaccades is influenced by top-down processes (Betta & Turatto, 2006, Otero-Millan et al., 2008).

These movements are what constitute what we understand as fixations. These sequences of fixations are important to understand how a scene is processed, considering that it is in these moments that information about that scenario is collected. As such, eye-tracking can be extremely helpful in understanding strategies. Therefore, it is necessary to understand what applications exist with eye-tracking for the strategies.

In the next chapters, concepts related to the use of eye-tracking in matrix reasoning and reading tests will be presented. In chapter 3, a theoretical review of cognitive strategies in matrix reasoning tests will be presented. In chapter 4, a study relating cognitive strategies in matrix testing with cognitive skills will be presented. In chapter 5, a study of cognitive strategies in a test of reading with gaps will be presented. Finally, in chapter 6, a general conclusion of all studies will be presented.

CHAPTER 3: COGNITIVE STRATEGIES IN MATRIX REASONING TASKS: STATE OF THE ART[1]

Introduction

Fluid intelligence (Cattell, 1963), or fluid reasoning (Gf; Carroll, 1993), can be defined as a set of abilities that together form a broad ability that have the main objective of solving new problems (Schneider & McGrew, 2012). In this way, abilities such as deliberate control of attention are employed in tasks that cannot be solved based on schemes. This type of intelligence is evident in abstract reasoning, which has less dependence on previously acquired knowledge. Therefore, Gf is explicit in inferential reasoning, relational reasoning, classification of new events, generalization and application of old concepts and solutions in new contexts and problems, creation of hypotheses, in addition to the identification of similarities or differences (McGrew, 2009). There is also a strong relationship between Gf and the perception of consequences from new knowledge and extrapolation of concepts (Schneider & McGrew, 2012).

To evaluate Gf, there are several types of tests (for a description of several tests, see the third section of Flanagan & Harrison, 2012). One of these tests is the matrices reasoning tasks. These tests are based on a matrix, which is most often a 3x3 matrix, where a part of it is missing and a set of alternatives to complete the missing part of the matrix (see Figure 3.1). Of course, a matrix reasoning task assesses matrix reasoning, but in addition, it also assesses inductive reasoning, deductive reasoning, spatial visualization, and non-verbal reasoning (Drodrick et al., 2012). The most well-known example of this type of test is the Raven’s Progressive Matrices (Raven et al., 1998), but there are other tests, such as *Wiener Matrizen-Test-2* (WMT-2; Schlottfeldt & Malloy-Diniz, 2018).

[1] The text in this chapter is based on a paper, currently in peer review, coauthored by the author and Elizeu Coutinho de Macedo.

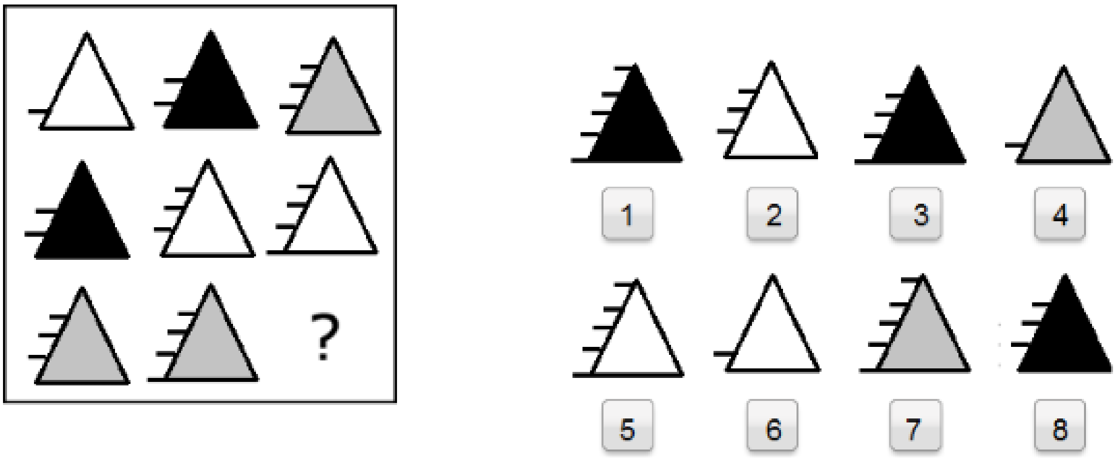

Figure 3.1 – Example of a matrix reasoning task item. The left structure is the 3x3 matrix with a missing piece. In the right, possible response alternatives to complete the matrix. The item is based on a WMT-2 item.

Matrix reasoning tests normally use the number of correct items and sometimes the total testing time to access the participant's score (Raven et al., 1998; Schlottfeldt & Malloy-Diniz, 2018), but this type of data gives little access to the cognitive processes of the participant during this task. One of the only ways of extracting cognitive processes with these measures is to see if the participant chose an answer that has one of the variables correct. In the example in Figure 3.1, the correct answer is the number 1, but if the participant chose items 3 or 8, it would be possible to say that they understand that the answer had to be black. Similarly, if the participant chose item 5 or 7, it would be possible to affirm that the participant understands that the item had to present 5 horizontal lines in the left. Therefore, cognitive processes can be analyzed by looking at errors, but since the common practice is to only get the numbers of correct answers, all this cognitive process is lost and rarely analyzed by researchers.

The analysis of errors can present some cognitive processes but not all since there are strategic processes that cannot be accessed by this type of analysis. The lack of information generated by these measures creates a need for new measures. Thus, several methods can be employed to measure it, such as eye-tracking (e.g., Vigneau et al., 2006), mouse-tracking (e.g., Rivollier et al., 2020), self-report questionnaires (e.g., Jastrzębski et al., 2018; Mitchum & Kelley, 2010), and think-aloud protocols (e.g., Jarosz et al., 2019). With these methods it is possible to infer some of the cognitive strategies used by the participants thorough their eye movements and use it for how the strategies affect the performance in the reasoning tasks (Laurence et al., 2018; Vigneau et al., 2006); and how the strategies change in populations with developmental disabilities (Curie et al., 2016; Vakil & Lifshitz-Zahavi, 2012). Furthermore, these methods can also be applied to the relation between working memory capacity, fluid reasoning and the strategies people adopt (Gonthier & Roulin, 2020; Gonthier & Thomassin, 2015; Jarosz & Wiley, 2012; Jarosz et al., 2019; Jastrzębski et al., 2018); how the strategies can interfere in the matrix reasoning task, with a look more into the test itself (Arendasy & Sommer, 2013; Becker et al., 2016; Mitchum & Kelley, 2010); and how the strategies can affect the test score of the participants (Hayes et al., 2015; Loesche et al., 2015).

Keeping in mind that there is a lot of research being done on cognitive strategies on matrix reasoning task, beginning in 2006 with the paper by Vigneau, Caissie, and Bors, it is surprising that no reviews have been done in this topic. Therefore, it is necessary to summarize all the advances that had been done in the last decade and a half on cognitive strategies on matrix reasoning tasks, as well to explain the problems that were found and future directions. This helps new researchers that are coming to this field and also presents the big picture for the researchers that are already doing their research on cognitive strategies in matrix reasoning tasks. Below, we provide the initial steps of the research that was made in this topic, the applications that were made, new methods that are emerging, and future directions.

Measuring strategies in matrix reasoning tasks

Snow (1980) proposed that there are two types of strategies that can be used in multiple-choice tests. The first one is the constructive matching strategy. This strategy is defined by the participant looking to the problem and finding an answer based on the patterns that are presented in the problem. Bethell-Fox and colleagues (1984) proposes that this strategy “seems to involve the construction of an idealized answer which is compared to the response alternatives” (p. 207). After idealizing the answer, the participant will search in the alternatives the answer that corresponds to the answer that they imagined. A representation of this type of strategy can be found in Figure 3.2a.

The other strategy proposed by Snow (1980) is response elimination. In this strategy, participants will try to eliminate the answers based on simple characteristics of the problem and then finding the correct one. Bethell-Fox and colleagues (1984) state that this strategy “involves a process of feature comparison between stem and alternatives aimed, presumably, at eliminating incorrect alternatives to arrive at the correct answer by default” (p. 207). This strategy’s representation is shown in Figure 3.2b.

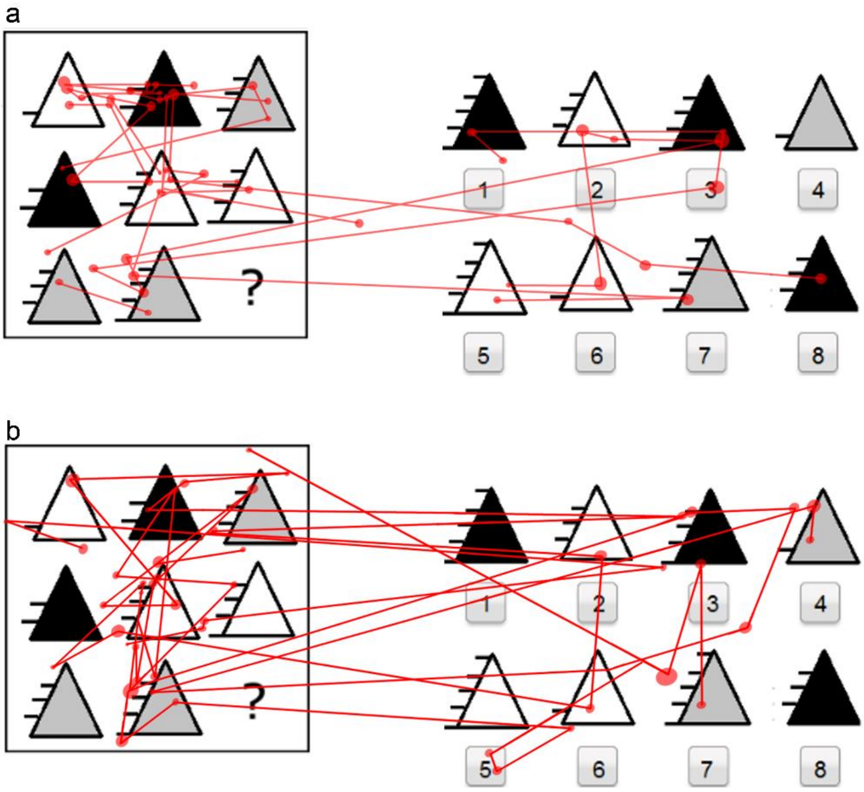

Figure 3.2 – Real eye movement patterns classified as the strategy of (a) constructive matching and (b) response elimination in a matrix intelligence test.

Bethell-Fox and Colleagues (1984) investigated these eye movements strategies in a geometric analogical task that had a sequence of figures regarding A is to B as C is to D, but the D was deleted. The participant had to fill the sequence with one of four alternatives that were displayed below. Their findings suggested, as Snow (1980) proposes, that participants that use the constructive matching strategy present higher abilities, while participants that use the response elimination strategy present lower abilities.

Vigneau and colleagues (2006) based their research paper on these findings and theories since the matrix reasoning tests are tests with multiple choices. To create a non-categorical measure, they developed a set of measures that could be used to determine which strategy a participant had employed. To do this, they created two types of measures that are calculated for each item of the test: measures of time on areas of interest; and measures of transitions (or alternations between interest areas, as the authors referred). Before we go on, it is important to note that we chose to change the name of some measures with the intention to reach a more precise language. The names provided by Vigneau and colleagues (2006) are provided inside parentheses.

The measures of time on areas of interest are basically three: the total time spent on the item (item latency); the time spent looking at the matrix (time-on-matrix, per the authors); and the time spent looking at the answers (time-on-alternatives, as the authors described). From these measures, several other can be calculated, such as: percentage of time looking at the matrix (proportional time-on-matrix, per the authors); percentage of time looking at the answers (proportional time-on-alternatives, as the authors wrote); matrix-answers ratio time. These measures are interesting because they equalize the time spent by the total amount of time in the item. For example, if participant A spent 30 seconds in the matrix and 10 seconds in the answers, while participant B spent 30 seconds in the matrix and 70 seconds in the answers, they would account the same time spent in the matrix. But if one corrects the measure by the total amount of time spent on the item; participant A would have spent 75% of the time on the matrix, while participant B would have spent 30% of the time on the matrix, indicating a big difference in the processing of the test. Furthermore, it is possible to divide the time spent on the matrix by the time spent on the answers, yielding a ratio between the two. In our example, participant A would have a ratio of 3, indicating that this participant spent 3 times more time on the matrix than on the answers, while participant B would have a ratio of 0.43, explaining that for every 1 second spent on the answers, the participant spent 0.43 seconds on the matrix.

The measures of transition are twofold: matrix-answer transitions (or number of toggles, as the authors wrote) and latency to the first fixation on an answer choice (Latency-to-First-Toggle, per the authors). Additionally, it is possible to calculate other measures, such as rate-of-toggling. Rate-of-toggling is calculated by getting the matrix-answer transitions for an item and dividing by this same item latency. This process will equalize the transitions by the time since a participant that stays more time in the item will naturally do more toggles (Laurence et al., 2018; Vigneau et al., 2006). In this sense, the rate-of-toggling will produce a measure of toggle per second. For example, a participant that does 5 toggles during a trial and stayed in this trial for 20 seconds will have a rate-of-toggling of 0.25 toggles per second. This measure is not so natural to understand; a more natural and understandable approach would be to see how many seconds the person spent toggling each time. In our example, this number can be achieved by dividing the seconds by the toggles, so we would get a final result of 4. Therefore, it would be possible to say that the participant made one toggle every 4 seconds, on average, making our measure much more understandable.

These measures are interesting because it is possible to use them to do approximations about which strategy the participant used. For example, as Vigneau and colleagues (2006) described, a higher number of measures of transition, such as matrix-answer transition or rate-of-toggling, can be related to the response elimination, while a lower number of measures of transition can be related to a constructive matching strategy. This happens because, in the response elimination strategy, the participant will make lots of transitions while trying to eliminate some of the answer choices. This pattern reflects the number of matrix-answer transitions. Furthermore, a higher latency to the first fixation on an answer choice is related to the constructive matching, since technically the participant would investigate the matrix longer before moving to the answers. Lower latency to the first fixation on an answer choice would indicate that the participant moved quickly to the answer choices, probably presenting a response elimination strategy. This is also clear in the measures of time in interest areas. For instance, more time spent looking at the matrix or proportional time spent looking at the matrix relative to the answer choices should be related to a constructive matching strategy, whereas the reverse can be related to a response elimination strategy. Likewise for the time spent looking at the answers or the proportional time spent looking at the alternatives: a higher number can indicate a response elimination strategy while the contrary could indicate a constructive matching strategy. These phenomena may happen because a participant that is doing a constructive matching strategy will focus primarily on the matrix, trying to create an idealization of the answer, while a participant adopting response elimination strategy will alternate several times between the answer choices, increasing their time in that area of interest (AOI).

One other way to investigate cognitive strategies is to use mouse-tracking. Rivollier and colleagues (2020) used this method but had to adapt the task in order for it to be a good measure. The authors created an interface that omits one of the two parts of the task. The task is divided into two parts: the matrix and the alternatives. To be able to see one of the parts, the user should move the mouse to that part. For example, while the mouse is on top of the matrix, only the matrix is shown to the participant. When the participant moves the mouse to the answer choices area, the matrix is hidden and then the answer choices appear to the participant. With this dynamic in the interface, it is possible to calculate the same measures that can be extracted from eye-tracking, such as time in each part of the task and the number of transitions. In this case, they are measured by the time that the mouse stayed on top of each part of the task and how many times the mouse went from a part of the test to the other.

One interesting aspect of the measures of transition and time on a part of the task is that, although they were created based on the use of eye-tracking, it is possible to investigate it without eye (or mouse) tracking. Mitchum and Kelley (2010) create a post-experiment questionnaire to access which strategy the participant used. Basically, participants were asked after the test, which of the potential strategies they most used: "(a) Looked at each response choice until I found one that seemed to fit; (b) Tried to predict what the correct answer should be and then searched for it among the response options given; (c) A little of both; or (d) Other" (Mitchum & Kelley, 2010, p. 701). If the participant chose option (c) or (d), they were asked to give further details. Gonthier and Thomassin (2015) made improvements in the questionnaire of Mitchum and Kelley (2010). Gonthier and Thomassin (2015) used a Likert-style scale with 9 levels (1: not at all true; 9: completely true) for two questions accessing constructive matching strategy (e.g., "I took the time to examine the drawing and to think about the answer before examining the response alternatives"; "After examining the drawing, I imagined the missing piece and then looked for it among the possible answers") and two questions regarding response elimination strategy (e.g., "After examining the drawing, I ruled out the response alternatives that did not match until only one remained"; "I successively examined each possible answer to decide whether it could be the missing piece").

Jastrzębski and colleagues (2018) made some additions to Gonthier and Thomassin (2015) questionnaire. They used an 8-point Liker scale (1: never; 8: always) and added two more items for each strategy (e.g., constructive matching: "First, I tried to understand how the correct answer should look like, and only at the end I looked at the answers"; "I spend most of the time trying to understand the problem without analyzing the answers"; response elimination: "When analyzing the problem I was simultaneously analyzing the answers", and "When solving a problem I spend most of the time analyzing the answers").

Another protocol to access the strategy is the think-aloud protocol. Jarosz and colleagues (2019) applied this protocol and asked participants to talk aloud what they were doing during each test item. In this sense, participants should talk aloud their thoughts on what they are doing in the test. Before going to the intelligence matrix test, they did two warm-up tasks. Afterwards, they did the matrix reasoning task, being videotaped. This method can decrease the percentage of correct answers in the test (Raven et al., 1998). The classification of the strategy was based on this recording and some the examples of the phrases can be viewed in Jarosz and colleagues (2019, p. 6). Jarosz and colleagues (2019) also identified a third strategy, called "isolate-and-eliminate", which is a hybrid of constructive matching and response elimination. This strategy is based on identifying one of the rules that govern the test item, eliminate the possible answers, and then trying to find another rule until one item is left.

In addition to the measures presented until now, which analyze the whole task trial (i.e., the matrix and the answer choices), there are other measures that can be used to analyze a specific part of the trial. Vigneau and colleagues (2006) presented an innovative measure of time distribution across the matrix cells called matrix time distribution index (MTDI). As Vigneau and Colleagues (2006) proposed, "this index reflects the distinction between partial and full matrix analysis" (p. 266). They suspected that if the participant focuses more in the rightmost column and in the line further down, the only column and line that have the missing piece, then in the whole matrix, then he is doing a partial analysis of the matrix. This would happen because if the participant only focuses on the rightmost column and on the line further down, it meant that he is trying to understand the missing piece only, instead of trying to understand the logical pattern presented in the matrix. To convert this hypothesis to an eye-tracking measure, the MTDI was calculated by the sum of the time spent on the rightmost column and on the line further down and then subtracting this value from the time spent on the other part of the matrix that was not the rightmost column and on the line further down. Therefore, if the MTDI is zero or a positive number, it means that the participant had a full matrix analysis. On the contrary, if the MTDI is a negative number, it means that the participant had a partial analysis of the matrix.

These were the first steps to understand the strategy use in matrix reasoning tests. From what Vigneau and colleagues (2006) created, several studies applied this in different areas and came upon novel information.

Application

Several studies applied this method to investigate different questions. Probably the most obvious question is how these strategies affect the performance on intelligence tests. In this sense, some authors (Laurence et al., 2018; Vigneau et al., 2006) tried to understand this relationship. The first study regarding this relationship was the same that created the method. Vigneau and colleagues (2006) found significant positive relations between the Raven Advanced Progressive Matrices (RAPM) score and percentage of time looking at the matrix, latency to the first fixation at an answer choice, and MTDI. The authors also found a significant negative relation between RAPM score and percentage of time looking at the answer choices, matrix-answer transitions, and rate-of-toggling. They also created models to understand how much these strategies could predict RAPM scores. They created 4 models: the first model used MTDI and percentage of time looking at the matrix as predictors; the second model used MTDI and latency to the first fixation on an answer choice as predictors; the third model used MTDI, percentage of time looking at the matrix, and latency on easy items; the fourth model used as predictors MTDI, matrix-answer transitions on easy items, and latency on easy items. Each model predicted 32%, 30%, 49%, and 51% respectively. These findings demonstrated that the strategies applied in the matrix reasoning test can predict a lot of the variation in the score of the test employed.

Laurence and colleagues (2018) also investigated this question. These authors used the WMT-2 instead of the RAPM. WMT-2 is an interesting test for this type of study because it has clear logical patterns in each item and the items are always based on only two variables. For example, in Figure 3.1, the varying elements are (1) the color and (2) the number of traces on the left of the triangle. This is interesting because it is a possible control for the cognitive load in each item. The authors found significant positive correlations between the WMT-2 score and item latency, time looking at the matrix, and latency to the first fixation on an answer choice. They also found significant negative correlations between the test score and percentage of time looking at the answer choices, and rate-of-toggling. Rivollier and colleagues (2020) also found a similar result with mouse-tracking: rate of toggling, latency to the first toggling, and proportional time on matrix and alternatives were significant in a mixed model with all the measures. Laurence and colleagues (2018) also tried to predict the score of the test and the logical patterns item groups based on the eye-tracking measures. For the whole test, rate-of-toggling was the only predictor, predicting 46% of the variation on the WMT-2 score.

Regarding the logical pattern item groups, Laurence and colleagues (2018) found that the easier logical pattern was predicted by item latency, time looking at the matrix, percentage of time looking at the matrix, time looking at the answer choices, percentage of time looking at the answer choices, matrix-answer transitions, and rate-of-toggling. These measures were able to predict 39% of the WMT-2 score variation. Rate-of-toggling was also the only predictor of the medium and harder logical patterns, predicting 22% and 29%, respectively.

Since there were different predictors, Laurence and colleagues (2018) also tried to replicate the first two models found by Vigneau and colleagues (2006). The third and fourth models were not replicated because Vigneau and colleagues (2006) used sub-test measures (e.g., measures in easy items) but these measures could be less reliable due to the lower number of items in WMT-2. The first model of Vigneau and colleagues (2006) predicted 9% of Laurence and colleagues (2018) data, while the second model predicted 34%.

The findings by Laurence and colleagues (2018) indicates that MTDI is not a good predictor of the test score—a result that was also found by Hayes and colleagues (2011)—but it can be used in other ways to understand what happened in a matrix reasoning test. Furthermore, based on the results presented by Laurence and colleagues (2018), the best measure to predict the test score is the matrix-answer transitions. This measure indicates that fewer transitions are related to a better test score.

These findings are also important for developmental disabilities since it can show how different clinical populations process reasoning tests. Participants with non-specific intellectual disability and participants with Down syndrome tend to begin looking at the answers earlier - i.e., spend less time/fewer eye movements studying the matrix before looking at the answer choices - then neurotypical counterparts (Vakil & Lifshitz-Zahavi, 2012). Additionally, participants with non-specific intellectual disability and participants with Down syndrome tend to do more matrix-answer-transitions, indicating that the participants with these disabilities probably rely more on a response elimination strategy than typical developed participants. This probably happens because participants with these disabilities would analyze the matrix more superficially than their typically developed counterparts (Vakil & Lifshitz-Zahavi, 2012). Additionally, the eye-movement pattern between these developmental disabilities and typical development were easily distinguishable, but this distinction was not present between participants with non-specific intellectual disability and participants with Down syndrome (Vakil & Lifshitz-Zahavi, 2012).

Since the RAPM can be too hard for participants with a disability, another option is to adapt it for these specific populations. With this in mind, Curie and colleagues (2016) published a study where they adapted RAPM to these populations and investigated the eye-movements of participants with intellectual disabilities compared to matched controls by chronological age and mental age. The adaptation was based on creating a 2x2 matrix, instead of 3x3, and presenting only 2 possible answers. Similarly to the prior literature (Vakil & Lifshitz-Zahavi, 2012), they found that the participants with intellectual disabilities focused a greater percentage of their time on the answer choices compared to the matched controls, and a lower percentage of their time on the matrix compared to the matched controls (Curie et al., 2016). Furthermore, the latency to the first fixation on an answer choice also presented a significant difference between the controls and the intellectual disability group, with the latter having a much lower latency (Curie et al., 2016). Curie and colleagues (2016) also state that they found a response elimination strategy in the intellectual disabilities group because of the matrix-answers transitions.

As we mentioned before, the matrix-answer transitions can be very important measure to understand the role of working memory (WM) capacity (WMC) in a matrix reasoning test because these transitions can be related to how much a person can process and maintain the information in mind. Thus, several studies (e.g., Gonthier & Roulin, 2020; Gonthier & Thomassin, 2015; Jarosz & Wiley, 2012; Jarosz et al., 2019; Jastrzębski et al., 2018) explored the relation between WMC the strategy use, as well as the relation of both with the Gf. It is noteworthy that several other studies (e.g., Engle et al., 1999; Mogle et al., 2008; Unsworth et al., 2009) investigated the relation between WMC and Gf, but we focused only on the ones that also looked on the strategy use.

To the best of our knowledge, the first study to investigate this subject on the light of strategy use was of Jarosz and Wiley's (2012) experiment 2. In this study, the authors manipulated the answer choices of the tests and created two categories: the high and low salience categories. In the high salience category, participants were shown 4 answer choices, with one of them being the most commonly chosen wrong answer. The low salience category also presented only 4 answer choices, but the biggest distractor was one of the answers omitted. Their findings in the second experiment demonstrated that participants' performance on high salience items predicted WMC and that a higher rate-of-toggling was one of the predictors of low WMC. These results suggested that higher WMC participants do a constructive matching strategy most of the time.

Based on the findings of Jarosz and Wiley (2012), Gonthier and Thomassin (2015) decided to investigate if the strategies could mediate WMC and Gf. These authors based their hypothesis on the idea that participants with higher WMC are able to do several strategies, which can be more effective and could lead to better performance. They used questionnaires to assess strategy use and found that this measure fully mediated the relationship between WMC and the performance on the matrix reasoning test. Jarosz and colleagues (2019) also found similar results. A partial mediation between WMC and Gf, measured by a matrix reasoning test, was identified while using a think-aloud protocol to assess strategy (Jarosz et al., 2019). However, Jastrzębski and colleagues (2018) reached an opposing conclusion. With a bigger sample, they failed to find a mediation by the strategy use. Jarosz and colleagues (2019) argued that maybe one of the problems with Gonthier and Thomassin (2015) and Jastrzębski and colleagues (2018) studies are the way that they are measuring the strategy: they argued that the think-aloud protocol is more valid than a questionnaire.

Still on WMC, Gonthier and Roulin (2020) found that as the items get harder in a matrix reasoning task, it is possible to predict in which item the participant will change from constructive matching to response elimination based on their WMC. In other words, participants with higher WMC relied from beginning to end on the constructive matching strategy, while participants with lower WMC changed from the constructive matching strategy to the response elimination strategy as more difficult items were presented. These results are particularly interesting because they show how the participants shift their strategy as the test goes on. At the end of the test, in the more difficult items, the use of response elimination and constructive matching was similar between participants, but at the beginning of the test, participants relied much more on the constructive matching strategy, indicating that at easy items, participants use constructive matching but only a few can employ this strategy on difficult items. These results confirm the predictions of earlier studies (Bethell-Fox et al., 1984; Snow, 1980). The results of this study can also be understood in relation to how the strategies interfere in a matrix reasoning task.

Other studies also looked at how the strategies can interfere in the matrix reasoning task. These studies were more interested in question regarding the psychometrics of the test based on the information of the strategies (Arendasy & Sommer, 2013; Becker et al., 2016; Mitchum & Kelley, 2010). In this sense, constructive matching appears to boost the test-taker's confidence in their answers (Mitchum & Kelley, 2010). This may happen because when a participant does the constructive matching strategy, they create an answer in their mind and this gives the sense of understanding what is happening in the test. By manipulating the matrix reasoning task in order to make the participant think about the answer first, Mitchum and Kelley (2010) forced participants to use the constructive matching strategy and were able to show that the confidence and accuracy of participants were higher in this condition. Rivollier and colleagues (2020), using the mouse-tracking, also forced participants to use the constructive matching strategy by making them look only at the matrix and after some time allowing them to see the possible answers. In this case, proportional time on the matrix and answer choices had the best correlation with accuracy, with more time proportional time on matrix leading to better results.

Furthermore, Arendasy and Sommer (2013) demonstrated that the strategy can affect the construct validity of the matrix tests. They proposed a format for the answers that can lead to less response elimination strategy and would help with the validity. Becker and colleagues (2016) went further and demonstrated that if the participant had to create their own answer, based on some given individual elements (e.g., a line, triangle, etc.), this prevents the response elimination strategy and increased the validity.

This is also true for participants who are taught the rules of the test before the test. In this sense, participants who knew how the logical patterns of the test work engaged much more on the matrix, doing a constructive matching strategy, instead of checking the answers (Loesche et al., 2015). In the same direction, these training can elevate a participant's score, but these gains after training can be related to differences in the strategy. Namely, the change in strategy while doing a matrix reasoning task after training can account for one third of the gain (Hayes et al., 2015). In this experiment, Hayes and colleagues (2015) applied a new method to investigate strategy use. This new method (Hayes et al., 2011) is categorized in a new cluster of analysis. These procedures rely much more on complex algorithms and analyze the sequence of transition between the areas of interest. This is the subject of our next section.

Emerging methods

The analysis of strategies in matrix reasoning tasks is relatively new. The first method was described by Vigneau and colleagues (2006) and has been the most used method ever since. This method is relatively simple to compute in a spreadsheet, and the results are relatively easy to understand. But since then, two methods (e.g., Hayes et al., 2011; Kurcharsky et al., 2020) have emerged, using complex algorithms to compute the sequence of transitions between each one of the areas of interest over time. This is beneficial because eye-tracking data have a temporal aspect that is eliminated by methods such as the one proposed by Vigneau and colleagues (2006; Holmqvist et al., 2011). Both new methods that are presented here can be used as an exploratory method for the strategies applied in matrix reasoning tasks (Hayes et al., 2011; Kurcharsky et al., 2020).

The first method was the one presented by Hayes and colleagues (2011). This method is based on transitions matrix, association between events, and Markov models. This method creates a transition matrix. Since there are 9 AOIs in the matrix plus the answer choices (that counts as one AOI), the transition matrix is a 10x10 matrix. This matrix has the senders as the lines and the receivers as the columns. Consequently, it is possible to read the probability based on the Markov model: since a fixation occurred in one AOI, therefore there is a known probability. Based on the pattern of transitions in the transition matrix, the method will understand the transition that happened, update the transition probability, and then update all the expected transitions that will happen after it. Therefore, this method considers higher order transitions, such as second-order transitions. In other words, based on a set of sequential fixations, the method will provide the probability of where the next fixation will happen. As Hayes and collaborators (2011) mentioned "[t]his is equivalent to learning to predict future scanpaths based on past scanpaths" (p. 2). The method is very sophisticated and produces a complex result. Thus, its application is not simple, and few papers have used it (e.g., Hayes et al., 2015).

The second method is the one published by Kurcharsky and colleagues (2020). This method is not exclusive to matrix reasoning tasks. It has an unsupervised algorithm that finds clusters of eye-movements that are similar between each eye-movement in each trial. Therefore, it is not possible to dictate to the algorithm what to expect, it finds clustered eye-movements and the researcher has to categorize it. The algorithm is based on a transition matrix (Kurcharsky et al., 2020). This method is based on k-means clustering. The clustering is made for 1 to 10 clusters, and after that, the scree plot from *k*-means clustering will be produced. With the scree plot, it is possible to select the best cluster numbers, if the scree plot presents an elbow. If not, some theoretical assumptions can be made. In their paper, Kurcharsky and colleagues (2020) did not find an elbow and chose 2 clusters based on previous research. Another decision that the researcher must do is if fixations on the same AOI will be used or deleted. Generally speaking, if a latent strategy can be found with repeated fixations, then it should not be deleted. The other way around is also true. After this, the method will produce an average transition matrix based on the hidden Markov model, similar to the transition matrix of Hayes and colleagues (2011). This transition matrix can be visually analyzed by the researcher and categorized regarding each strategy.

After the clustering is done, each participant trial has a categorization in one of the clusters. For example, based on the transitions that a participant made on the trial, this trial will have a categorization that is based on similar pattern of transition of other participants. With this categorization, several analyses with more common statistical methods can be used. As an example, it is possible to use this clustering in a whole dataset of participants that went through different conditions and then analyze if the categorization differed for each condition. Therefore, this method is useful to categorize different datasets and produce new analyses on top of it.

Kurcharsky and colleagues' (2020) method was applied to the Mastermind game and on the data of Laurence and colleagues (2018) with WMT-2. In the matrix reasoning task, the algorithm was coded to gather the data in two clusters, since there was no "elbow" in the scree plot and past research supposed two clusters. The first one is similar to the constructive matching, but the second one was not related to the response elimination. Indeed, both strategies had something similar, but the algorithm did not categorize all eye gazes as one strategy (Kurcharsky et al., 2020). This is evidence against the model of strategy proposed by previous research (Bethell-Fox et al., 1984; Snow, 1980; Vigneau et al., 2006).

Kurcharsky and colleagues' (2020) method differs from Hayes and colleagues' (2011) method regarding order of transitions. Kurcarshky and colleagues' (2020) method only look for first order transitions (i.e., where the fixation will happen next based on where the fixation just happened) while Hayes and colleagues' (2011) method analyzes higher order transitions, in specific, second-order transitions (i.e., where the fixation will happen next based on where the last two fixations happened).

The possibility created by such methods is groundbreaking for research on cognitive strategies based on eye-movements. Since categorization is possible using these methods, new ways to understand the data are created. As an example, since a non-supervised categorization is possible, this can be tested with several cognitive measures in order to understand how each type of strategy is related to a cognitive trait. The questions asked by Gonthier and Roulin (2020), Gonthier and Thomassin (2015), and Jastrzębski and colleagues (2018) regarding WMC and the strategies can be revised with these algorithms. Furthermore, analyses such as the ones proposed by Vigneau and colleagues (2006) failed to acknowledge the process that each participant went through during the trials (i.e., the sequence of transitions made by the participant during the trial). With these new methods, it is possible to understand the step-by-step of each participant and how this relates to, for example, cognitive measures. Therefore, we highly suggest that new research is based on these methods. The potential for new discoveries using this method is very promising.

However, one possible problem with these new emerging methods is that due to the difficulty of implementation, few papers have used them. This issue creates complications for the validity of the methods and the understanding of the implementation of such algorithms, although they can bring more information than the method by Vigneau and colleagues (2006). As we mentioned, such methods can analyze the sequence of eye movements instead of only looking at a summary of the measures. It is noteworthy that Kurcharsky and colleagues (2020) have made their code (R scripts) and data freely available (https://osf.io/wvzs9/). It remains to be seen whether this method will be adopted by the scientific community.

Future directions

As we have seen, several steps have been made regarding strategies in matrix reasoning tasks. Most of the papers were based on the idea brought by Vigneau and colleagues (2006), but Kurcharsky and collaborators (2020) did not find the same strategies using a non-supervised method.

It is hard to understand why the response elimination strategy was not found by Kurcharsky and colleagues (2020). One possibility is that there is no pure response elimination strategy. Instead, they may have found a mixed strategy between constructive matching and response elimination, as Jarosz and colleagues (2019) reported. Perhaps the isolate-and-eliminate strategy is more common than a pure response elimination strategy, and therefore the categorization was incorrect. Since participants tend to shift from a constructive matching to a response elimination strategy as they are not able to effectively proceed with the constructive matching (Gonthier & Roulin, 2020), the strategy would not look exactly like response elimination, but more like a frustrated constructive matching strategy. This would reflect in the strategy with a pattern that looks similar to a constructive matching eye-movement pattern, a result that Kurcharsky and colleagues (2020) found.

This issue will need to be settled in future research, since it is the basis of the whole literature on the topic of matrix reasoning strategies. Future studies should look at whether people sometimes use a response elimination strategy, whether there is a third strategy (i.e., isolate-and-eliminate strategy, Jarosz et al., 2019), or whether the only strategies participants use are constructive matching and isolate-and-eliminate.

Another methodological point to be aware of is that several protocols have been employed, such as eye-tracking (e.g., Vigneau et al., 2006), mouse-tracking (e.g., Rivollier et al., 2020), self-report questionnaires (e.g., Jastrzębski et al., 2018; Mitchum & Kelley, 2010), and think-aloud protocols (e.g., Jarosz et al., 2019), but no study has attempted to examine the relationship between these methods. Thus, new studies should be conducted trying to understand whether there are clear relationships between these measures. For example, is there a strong correlation between the scores of the questionnaire presented by Jastrzębski and colleagues (2018) and the measures of eye-tracking created by Vigneau and colleagues (2006) or the new methods of Hayes and colleagues (2011) or Kurcharsky and colleagues (2020)? Rivollier and colleagues (2020) demonstrated that the mouse-tracking can present cleaner statistical results than eye-tracking which has some noise in the data, but does the strategy presented in participants using mouse-tracking relates to the eye-tracking strategies?

Nevertheless, it is also important to understand how cognitive strategies relate to other constructs. The relation between the strategies in the matrix reasoning tasks and WMC has been one of the topics that has been investigated the most, but the relationship is still not clear or even if there is any relation (Jastrzębski et al., 2018). Therefore, more research is needed in this topic. Questions focused on how the strategies relate to other constructs such as cognitive flexibility, problem-solving, memory, attention control, cognitive inhibition, etc., or if there is a clear relationship between the strategies and WM remain open.

Furthermore, another noteworthy point is that most of the data produced in this topic were conducted with university-related participants, with only two exceptions that investigated clinical populations (e.g., Curie et al., 2016; Vakil & Lifshitz-Zehavi, 2012). The population scope needs to be expanded. How do the strategies develop or change over childhood? As cognitive functioning declines in older age (Ratcliff et al., 2011), does this affect the cognitive strategies participants adopt? How are strategies in clinical populations that do not have an intellectual disability (e.g., dyslexia)? How do these strategies change as a function of education or socioeconomic status (an example of this type of question, relating eye-movements with socio-demographic, in another type of fluid reasoning task can be seen in Kasneci et al., 2021)? All these questions remain unsettled. Additionally, most studies investigating directly the strategies have had a small sample size, with a few exceptions (e.g., Gonthier & Roulin, 2020; Gonthier & Thomassin, 2015; Jastrzębski et al., 2018; see Supplementary File 1 for info of all papers), who had a large sample. The sample's size varied from 34 to 1890 participants. But it is noteworthy that eye-tracking papers had samples between 34 to 137 participants, with only one paper with a sample over 100 participants. Therefore, larger samples are needed to avoid underpowered studies in this field.

Lastly, Jastrzebski and colleagues (2020) mentioned that tasks such as matrix reasoning tasks are hard to study because of the complex aspect of it since it has a big and complex reasoning period. Therefore, these authors suggested going a step down and studying reasoning tasks that are simpler. This can be a promising idea to understand visuo-cognitive strategies related to reasoning. Several studies (Hessels et al., 2011; Starr et al., 2018; Thibaut & French, 2016; Thompson, 2021; Vakil et al., 2011; Vendetti et al., 2017) investigated the visual strategies in simpler reasoning tasks and have made progress in accessing strategies in simpler tasks. Therefore, it is important that in future research this idea is taken into consideration. There is a great need for a study that relates these complex cognitive processes associated with the strategy in complex tasks, such as a matrix reasoning task, with simple reasoning tasks.

Conclusion

In this review, we were able to expose the state of the art of research on cognitive strategies in matrix reasoning tasks. We covered how the strategies were conceived and how they can be studied, the applications that were made using these methods, as well as what piece of knowledge this application produced.

Summarizing, applications were found regarding: how the strategies affect the performance in the intelligence test (Laurence et al., 2018; Vigneau et al., 2006); how the strategies change in populations with developmental disabilities (Curie et al., 2016; Vakil & Lifshitz-Zahavi, 2012); the relation between working memory capacity, fluid reasoning, and strategy use (Gonthier & Roulin, 2020; Gonthier & Thomassin, 2015; Jarosz & Wiley, 2012; Jarosz et al., 2019; Jastrzębski et al., 2018); how the strategies can interfere in the matrix reasoning task, with a look more into the test itself (Arendasy & Sommer, 2013; Becker et al., 2016; Mitchum & Kelley, 2010); and how the strategies can affect the test score of the participants (Hayes et al., 2015; Loesche et al., 2015).

Additionally, we discussed new methods that are being applied in research with eye-movements and cognitive strategies in matrix reasoning tasks (e.g., Hayes et al., 2011; Kurcharsky et al., 2020). We strongly recommend that new papers that focus on cognitive strategies in matrix reasoning tasks adopt one of these new methods, since they are more sophisticated and can produce better results that are based on the sequence of the fixations instead of a summary of the fixations.

Finally, we were able to conceive where future research should focus. Several gaps were identified, namely: confirmation of the strategies described by Snow (1980) and Bethell-Fox and colleagues (1984); understanding the relationship between the different methods (e.g., eye-tracking, mouse-tracking, self-report questionnaires, think-aloud protocol); the relation between strategies applied in matrix reasoning tasks with other constructs, such as executive functions; and investigating these strategies in other populations that are not university-related participants.

We hope that this review can benefit researchers of the area since most of them have been doing sparse works with few connections. Furthermore, we also expect that this review brings more relevance to this matter considering only a few research groups have been engaging in it. Finally, we believe that this review will be of value for new researchers who want to study cognitive strategies in matrix reasoning tasks.

CHAPTER 4: UNDERLYING COGNITIVE ABILITIES IN THE USE OF STRATEGY IN MATRIX REASON TASK[2]

Introduction

Fluid reasoning (Gf; Carrol, 1993) is a set of abilities that helps us solve new problems (Schneider & McGrew, 2012). Gf is highly demanded when a person needs to do relational and inferential reasoning, classification of new situations and phenomena, formulation of hypothesis generalization, application of old schemas in new events and problems, and establishing similarities and differences (McGrew, 2009). Further, the relationship between Gf and Working Memory (WM) is well documented (e.g., Chuderski, 2013, Dehn, 2017, Kaufman, 2014).

Several tests can evaluate Gf (see the third section of Flanagan & Harrison, 2012). These tests normally present a pattern of figures and objects, and the test-taker must understand the rule that is guiding this pattern and then construct the answer or choose the correct answer in the answer choices bank (Alves, 2007, Schlottfeldt & Malloy-Diniz, 2018). One of the most common types of test used to measure Gf is the matrices reasoning task. This type of test assesses deductive, inductive, matrix, and non-verbal reasoning (Drodrick et al., 2012). Examples of this type of task are Raven's Progressive Matrices (Raven et al., 1998), the WAIS matrices (Wechler, 2004), and the Wiener Matrizen-Test-2 (WMT-2; Schlottfeldt & Malloy-Diniz, 2018).

When doing this type of test, a test-taker will use visual cognitive strategies to try to solve the task. Vigneau and colleagues (2006) described two strategies in matrix reasoning tasks, based on the work of Bethell-Fox and colleagues (1984), and Snow (1978, 1980) in analogical tasks with alternatives. The strategies can be defined as constructive matching – an effective but costly strategy where participants will try to solve the matrix by mentally constructing the missing piece before going to the answer choices to look for it – and response elimination – a less precise strategy where participants will do multiple alternances between the matrix and the answer choices trying to eliminate wrong answer choices.

[2] The text in this chapter is based on a paper coauthored by the author, Tatiana Abraão Jana, Silvia A. Bunge, and Elizeu Coutinho de Macedo.

These strategies can be studied through the use of eye-tracking. For example, the number of times that a participant goes from the matrix to answer choices and vice-versa can be indicative of the strategy that they are using: more alternances can be related to a response elimination strategy while fewer alternances can be related to constructive matching. The same goes for how long it takes for a participant to go to the answer choices for the first time, or how much more time the participant spent on the matrix instead of the answer choices (see Laurence et al., 2018). Further, it is possible to use algorithms that classify the scanpaths of different participants based on their similarities. In this case, it is possible to use transition matrices on the areas of the matrix reasoning task and gather information of how participants proceed in the visual scan of the task and analyze this transition matrix to understand which strategy was used by the participants (Kucharský et al., 2020).

These strategy measures have a relation with the performance. Previous studies demonstrated that these measures could predict the performance in the Gf test, demonstrating that participants with higher Gf tend to do eye-movement behaviors similar to the constructive matching strategies (Laurence et al., 2018, Vigneau et al., 2006). This finding is somewhat obvious since different patterns of behavior in a task can change the outcome. However, we found no studies studying the relation of the eye-tracking measures with a Gf score that was not measured in the same test where the strategy where measured.

The strategies and their relationship with performance were also investigated taking the WM into account. In this case, several studies investigated the relationship of both strategies with WM (e.g., Gonthier & Roulin, 2020, Jarosz et al., 2019). These studies demonstrated a relationship where more WM is related to using the constructive matching strategy. In this case, participants tended to use constructive matching in easy items and turn to response elimination in hard items. Also, the WM capacity predicted the strategy use, with participants that presented higher WM capacity maintaining the use of constructive matching strategy even in hard items (Gonthier & Roulin, 2020). It is noteworthy that, although this relationship between the type of strategy used and WM is very common in the literature, no study proposed to investigate this relationship with aid of eye-tracking. Most studies used questionnaires about strategy use (e.g., Gonthier & Roulin, 2020) or verbal protocols (e.g., Jarosz et al., 2019).

Since WM is related to the retention and manipulation of information, it is a limiting ability for most test-takers (Gonthier & Roulin, 2020). Further, WM is related to cognitive flexibility. On one hand, dopaminergic input in the prefrontal cortex – a cortex region related to maintaining and relating information – will bring stability to the WM, blocking interference from distractors.

On the other hand, dopaminergic modulation of the striatum will generate cognitive flexibility, allowing new information to enter the WM (D'Esposito & Postle, 2015). This mechanism also relates to planning since these abilities are crucial for following a plan. In fact, the cortex regions related to WM also are evident when humans are planning (e.g., Anderson et al., 2005, Lindner et al., 2010). Lastly, all these cognitive abilities relate to executive function (EF). The EF are a set of abilities that allow us to complete tasks (Miyake & Friedman, 2012). Although these cognitive abilities such as cognitive flexibility, planning, and EF seem like an important component for the use of different eye-movement strategies, to the best of our knowledge, no study tried to study these relations, especially with the use of eye-tracking.

Fixation durations are also an eye behavior that is related to different cognitive abilities. First, individuals with high WM tend to do longer fixations (Luke et al., 2018, Meghanathan et al., 2015). Second, Hodgson and colleagues (2000) found that participants that had a good performance in the Tower of London test, a planning task, had less fixation time during the task. This indicates that good planners are capable of doing fast and efficient fixations. To the best of our knowledge, no studies verified the relationship of the fixations pattern and behaviors in a matrix reasoning task with different cognitive abilities.

With these literature gaps in mind, our study aimed to explore the relationship between cognitive abilities and eye-tracking measures related to strategy use in matrix reasoning tasks. To do this, we conducted two studies. In the first study, we measured Gf in a matrix reasoning task and a non-verbal reasoning task. In the matrix reasoning task, we recorded the eye movements of the participant and calculated measures regarding cognitive strategy use. In the second study, we measured planning, WM, cognitive flexibility, self-reported EF, and also Gf with a matrix reasoning task. In the matrix reasoning task, we computed the strategy use measures based on eye movements. We employed golden standard methods of machine learning for small samples by using train/test split (see Vabalas et al., 2019) in order to select eye-movement predictors for Gf, planning, WM, cognitive flexibility, and self-reported EF.

We defined hypotheses regarding the relation of the eye gaze metrics and the underlying cognitive processes. The hypotheses were pre-registered in the AsPredicted under number 45682 (https://aspredicted.org/LWU_FLY). The hypotheses are:

Hy1: We predicted that participants who perform better on planning would be more likely to adopt a constructive approach, trying to solve the matrix problem before looking at the answer choices. This would manifest as a higher ratio of time spent on the matrix vs. the answer choices.

Hy2: We predicted that better spatial working memory would correlate with several gaze metrics, but that the distinguishing characteristic (as compared with the other cognitive measures) would be fewer gaze transitions between the matrix and the answer choices. On this view, participants with better spatial working memory would be able to keep in mind what the answer should look like when they transition from the matrix to the answer choices -- or, conversely, keep in mind an answer choice and check whether it fits.

Hy3: We predicted that participants who make fewer perseverative errors on Wisconsin Card Sorting Test would revisit incorrect answer choices less - that is, that they would make fewer fixations on the incorrect answer choices.

Study 1

Methods

Participants

A total of 62 university students (40 women, 66.12%, Medianage = 21, Rangeage = 18-29) were recruited for this experiment, as part of a larger project. Two participants had an exceptionally low number of fixations detected (two or fewer fixations per trial). Thus, both participants were removed by low number of fixations. The remaining participants (N = 60, 39 women, 65%) were all university students and their age (M = 21.48, SD = 2.50) ranged from 18 to 29 years old. Data collection took place over three time periods: between April and October of 2016; between November of 2017 and April of 2018; and between November and December of 2019.

Instruments

WMT-2 (Eye-tracking task)

WMT-2 is a Gf matrix reasoning test similar to Raven’s Progressive Matrices. It has a total of 21 problems, with 3 being examples that don’t count to the final score (and were not analyzed) and 18 real problems. Each problem is composed of a 3x3 matrix, that is the problem, in the left and 8 alternatives in a 2 x 4 matrix in the right of the screen. We used the computerized version of the test (Schlottfeldt & Malloy-Diniz, 2018). Between each test screen, a black fixation point was presented in a gray background for 2 seconds.

D.70 Test

The D.70 test is a non-verbal inductive reasoning test. The test consists of 44 items that are a sequence of domino pieces with a missing piece. The participant’s objective is to write the correct number of dots in each domino cell based on the pattern in the number of the other pieces. Participants were given 25 minutes to try to solve all the items. We used the Brazilian version of D.70, in the paper and pencil format (Alves, 2007).

Apparatus

To record the eye-gaze data we used a RED500 eye-tracking from SensoMotoric Instruments. We sampled at a temporal resolution of 500 Hz. We used the iView^{TM} software (v. 3.7, SensoMotoric Instruments, Inc.) to calibrate the eye-tracking device and to record the data, the Experiment Center^{TM} (v. 3.7, SensoMotoric Instruments, Inc.) to present the stimuli, and used the BeGaze^{TM} software (v. 3.7, SensoMotoric Instruments, Inc.) to extract the data. We used the default calibration procedure with 9-points. The fixation and eye-data algorithm used was the default of BeGaze^{TM} with minimum fixations of 100 milliseconds.

Procedure

The study was approved by the University Ethics committee (CAAE: 75035917.5.0000.0084). Participants were taken to the experiment room and seated at a desk. The experiment was explained to them and if they agreed with their collaboration, they would give their written consent. They firstly answered the D.70 in the paper and pencil format. After finishing it, they were placed in a chair ~70 cm away from a computer screen with a diagonal of 22 inches. The calibration procedure was conducted. Participants were presented to the instruction screen of WMT-2, told how the test works, and had an opportunity to ask questions about it before beginning the experiment. Each test problem had 1366 × 786 pixels. When the participants had an answer to the problem, they were asked to provide their answer verbally, indicating the number corresponding to one of the answer choices aloud and the experimenter would skip to the next trial and register their answer. Upon completion of the study, participants received course credit.

Eye-tracking measures

The average eye-tracking ratio was of 95.3% (SD = 4.13), with the participant that had the lower tracking ratio having their eyes detect by 76.9% of the task while the participant with most tracking ratio presenting a tracking ratio of 99.8%. We excluded the first fixation in each trial and all the fixations that were not in the matrix or answer choices. Additionally, we only used fixations with a duration over 100 ms since we were interested in cognitive fixations (Pieters & Wedel, 2012). We calculated several eye-tracking metrics, based on previous research (e.g., Laurence et al., 2018, Vigneau et al., 2006), also created new variables, and calculated common eye-tracking measures (see Table 4.1 for the complete description of each variable).

We also used Kucharský and colleagues' (2020) method to classify scanpaths. This non-supervised method will calculate a transition matrix based on the AOIs for each scanpath in each trial and use the standard k-means clustering, an algorithm based on lowering the within-cluster sum of squared Euclidean distances, to classify each scanpath into k clusters based on their Euclidian proximity. The k will be the number of centroids used to classify each scanpath. Based on the literature (e.g., Kucharský et al., 2020, Vigneau et al., 2006), we classified each scanpath based into two centroids, following the idea of constructive matching and response elimination strategies. In other words, the scanpaths were classified into two possible non-supervised clusters. Furthermore, following previous research (Hayes et al., 2011, Kucharský et al., 2020), we opted to delete repeated fixations in the same AOI. To compare both clusters, we calculated the Bayes Factor for each eye-tracking measure presented in Table 4.1. The BF10 is the Bayes Factor giving the evidence for H1 over H0. By convention, values over 3 are considered moderate evidence in favor of the H1, values over 10 are considered strong evidence, and values over 100 are considered extreme evidence. On the other side, values under 0.33 are considered moderate evidence in favor of H0, values under 0.10 are considered strong evidence, and values under 0.01 are considered extreme evidence (Jeffreys, 1961).

Table 4.1 – Eye gaze measures used in this study and their definition.

Eye-tracking Measures	**Definition**
Number of matrix-matrix transitions	Number of times that a participant gazed from a matrix cell to another matrix cell.
Number of matrix-answer transitions	Number of times that a participant gazed from the matrix to the answer choices or vice-versa.
Number of answer-answer transitions	Number of times that a participant gazed from an answer choice to another answer choice.
Latency to the first fixation on an answer choice	The time it took for a participant to do the first fixation on the answer choices.
Ratio of time spent on the matrix vs answer choices	The time spent on the matrix divided by the time spent on the answer choices.
Average number of visits to a given matrix cell	The number of visits in each cell in all test screens and then calculating the mean of all visits.
Average number of visits to a given incorrect answer choice	The number of visits in each answer choice, excluding the correct choice, in all test screens and then calculating the mean of all visits.
Total number of fixations on matrix cells	The sum of the fixation count that a participant had over the Matrix AOI in the whole test.

Total number of fixations on answer choices	The sum of the fixation count that a participant had over the answer choices in the whole test.
Average fixation duration for a matrix cell	The sum of time spent fixating in matrix cells divided by the total number of fixations on matrix cells.
Average fixation duration for an answer choice	The sum of time spent fixating in answer choices divided by the total number of fixations on answer choices.
Percent of trials classified as cluster 2 scanpath	We used a non-supervised algorithm to classify the scanpath of each individual in each trial into two possible clusters. Next, we calculated the percentage of trials that a scanpath was classified into the second cluster for each individual.

Data Analysis

We used a Least Absolute Shrinkage and Selection Operator (LASSO) regression model to analyze which eye-tracking measure predicted the D.70 test score. LASSO regression carries out the L1 regularization in the predictors by employing a penalty (λ) to the coefficients. This relationship of the coefficients and the shrinking parameter can be represented as:

$$||y - x\beta||_2^2 + \lambda||\beta||_1$$

With this feature, coefficients that have low predictive power are penalized until hit zero, while coefficients with high predictive power are penalized but are not set to zero. In our study, this method will remove the eye-movement measures that are weakly associated with the cognitive measure. The shrinking parameter is user-selected, and, because of it, it is necessary to do cross-validation with several values of λ. To do it, we split our data into two parts: ~80% of the data (50 participants) was used to train the model and select the best penalty value based on the root-mean-square error (RMSE) in a Leave one out Cross-Validation; ~20% of the data (10 participants) was used to test the best model selected in the training dataset. To evaluate the model in the test set, we calculated performance estimates such as correlation coefficient (r); R^2; mean absolute error (MAE); and RMSE. The R^2 presents the explained variation by the model, while MAE and RMSE represent a measure of the error of the model. Both MAE and RMSE will tell how much error there is on the predictions, but MAE will tell the mean error of the model, while the RMSE will also be penalized by outlier errors. Since the measures of the model are in z-score, the MAE and the RMSE present errors in standard deviations.

Results

Descriptives

The descriptive of each variable used in this study can be found in Figure 4.1. The participants presented a mean (SD) score of 27.90 (4.70) in the D.70 test. The scores ranged from 18 correct answers to 36 correct answers. In the WMT-2, the sample mean score was 10.43 (3.02), ranging from 5 to 17 correct answers. Scores on the D.70 and WMT-2 tests were correlated, $r = 0.53$, $p < 0.001$.

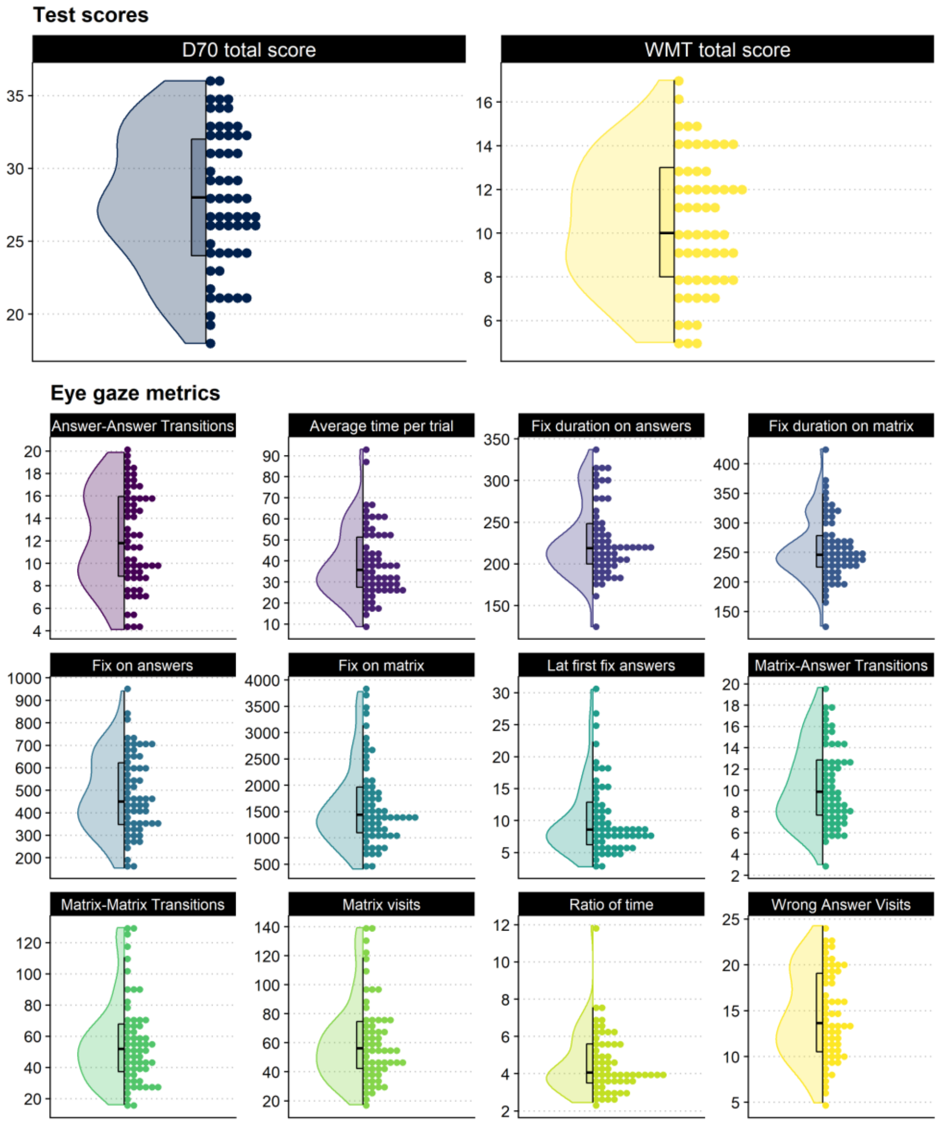

Figure 4.1 – Raincloud plots for the variables used in this Study 1.

The scanpaths of each participant in each trial were submitted to a classification analysis to identify different clusters of eye-movement strategies. In this analysis, two distinct – although similar – strategies can be identified. The first cluster presents a strategy pattern where participants explored the matrix row-by-row. In the second cluster, participants explored the matrix row-by-row, but also in a column-by-column pattern. The similarities and differences between the clusters are presented in Table 4.2. The transition matrix plots and the distribution of eye-movement measure plots for each cluster can be found in supplementary file.

Table 4.2 – Comparison of both non-supervisioned clusters.

Measures	**Cluster 1**	**Cluster 2**	**Analysis**
Gaze direction	Row-wise	Row-wise & Column-wise	Visual inspection (see supplementary file)
Probability to go to the answers when in top and mid row	Low	Low to moderate	Visual inspection (see supplementary file)
Average time spent in each trial	-	-	BF_{10} = 0.07 (±<0.00) ^{oo}
Matrix-matrix transitions	-	-	BF_{10} = 0.21 (±<0.00) ^{o}
Matrix-answer transitions	-	-	BF_{10} = 0.08 (±<0.00) ^{oo}
Answer-answer transition	More transitions	Less transitions	BF_{10} = 10.19 (±<0.00)**
Latency to the first fix. on an answer choice	-	-	BF_{10} = 1.09 (±<0.00)
Ratio of time	-	-	BF_{10} = 0.07 (±<0.00) ^{oo}
Visits to a given matrix cell	-	-	BF_{10} = 0.17 (±<0.00) ^{o}
Visits to a given incorrect answer choice	-	-	BF_{10} = 0.46 (±<0.00)
Fixations on matrix cells	-	-	BF_{10} = 0.07 (±<0.00) ^{oo}
Fixations on answer choices	-	-	BF_{10} = 0.38 (±<0.00)
Average fixation duration for a matrix cell	Longer fixations	Shorter fixations	BF_{10} > 1000 (±<0.00)***
Average fixation duration for an answer choice	Longer fixations	Shorter fixations	BF_{10} > 1000 (±<0.00)***

**Strong evidence for H1, ***Extreme evidence for H1, ^{o}Moderate evidence for H0, ^{oo}Strong evidence for H0.

LASSO regression model

Our LASSO regression model, predicting the D.70 performance from the eye tracking metrics (shown in Table 4.1), performed poorly, as judged by the performance estimates of the model (Table 4.3). Only two variables were selected as contributing significantly to model prediction: the number of matrix-answer transitions and the ratio of time spent on the matrix vs. answer choices. These variables were able to predict 6% of the variations of the D.70.

Table 4.3 – Coefficients and measures of the LASSO regression model predicting the D.70 total score.

Measures	Coefficients
Predictor's coefficients [a]	
Matrix-answer transitions	-0.16
Ratio of time	0.15
Performance estimates	
Correlation coefficient	0.25
MAE	0.72
RMSE	0.90

[a] Showing the predictors selected by the LASSO model; see full set of eye gaze metrics in Table 4.1.

Study 2

Since the eye-tracking measures were not heavily related to the Gf score, it is possible to wonder if the eye-movement measures are related with other cognitive measures. Therefore, we conducted a second study, with a new sample, in order to understand the relation of the eye-tracking measures in the matrix reasoning task with other cognitive measures such as planning, working memory, cognitive flexibility, and self-reported EF. This study was pre-registered in the AsPredicted website under registration number 45682 and can be accessed through https://aspredicted.org/LWU_FLY.

Method

Participants

We recruited 73 participants (47 women, 64.38%, Median_{age} = 21, Range_{age} = 18-33) for this study. However, 4 participants were excluded from the sample: three for poor eye-tracking data quality and one for low score in the Tower of London test. Therefore, our final sample had 69 participants (45 women, 65.2%). Our sample had a mean age of 22.46 years (SD = 3.49), ranging from 18 to 33 years. All the participants were university students. Most of them were coursing the law school (N = 31, 44.9%), followed by psychology (N = 23, 33.3%). Other students (N = 15, 21.7%) were majoring in engineering, physical therapy, pharmacy, architecture, economy, journalism, or neuroscience. The data collection occurred between February and November of 2017. Contact was made with the participants through social networks linked to the university and through the snowball sampling method.

Instruments

WMT-2 (Eye-tracking task)

The same test described in study 1. The eye gaze of the participants was recorded during this task.

Tower of London test

The Tower of London test is a test that assesses planning ability and logical reasoning. The test has 12 target figures that participants should try to reach. At each level, the difficulty to reach the target figure increases. For each target figure, three attempts are allowed, and the answer is only considered correct if the solution is reached in the correct number of allowed moves. Thus, the score for each level ranges from 0 to 3, depending on how many times the participant has tried, and the total score ranges from 0 to 36 (Shallice, 1982, Krikorian et al., 1994).

Corsi block-tapping test

The Corsi block-tapping test is a visuospatial working memory test. The test consists of two parts: in the first, the researcher will touch the blocks in a sequence and the participant will have to repeat the same movements in that sequence; in the second part, the evaluator will touch the blocks in a sequence and the participant will have to repeat the sequence inversely. The difficulty of the sequence increases with every two sequences made, with one more touch being added to the sequence. The test ends when the participant misses two sequences with the same number of touches. The total number of sequences is 14. The total test score varies between 0 and 28 (Corsi, 1973. Santos et al., 2005). The Corsi block test presents evidence that it can be used in the Brazilian environment (Santos et al., 2005).

Wisconsin Card Sorting test (WCST)

The Wisconsin Card Sorting Test is a test of abstract reasoning, the use of cognitive strategies in changeable environments, and cognitive flexibility. It is a test that is closely linked to executive functions. In this test, the participant is presented with a sequence of 128 cards and must speak to which categorization criteria they are grouped. Criteria can be color, shape, or number of stimuli. Criteria will change after 10 hits in a row. This test can be evaluated by different types of measures, but we used perseverative errors, a measure of inhibitory control, and cognitive flexibility. The test was adapted to the Brazilian context and can be used in this population (HEATON et al., 2004).

Behavior Rating Inventory of Executive Function for Adults (BRIEF-A)

The BRIEF-A is a self-report questionnaire. It assesses self-regulation in daily life on adults aged 18 to 90 years. It is 75 Likert-type items with three levels: “never”; "sometimes"; "often".

Item score is as follows: never = 0; sometimes = 1; often = 2. The items present statements such as “I have trouble with jobs or tasks that have more than one step” or “I make mistakes carelessly” and were created based on executive function concepts. The total BRIEF-A score is given by the sum of the behavioral regulation and metacognition indices. A higher score on this scale indicates executive dysfunction (ROTH et al., 2005; ROTH et al., 2013). For the present study, the translation made by Jana (2018) was used.

Apparatus

The same used in study 1. The data was recorded at a temporal resolution of 500 Hz.

Procedure

The project was submitted to the Ethics and Research Committee and approved under CAAE number 63883016.0.0000.5487. Data collection was performed in a single session in the laboratory. Upon arriving at the laboratory, participants received explanations about the study and, if they wanted to continue with the participation, signed two copies of the Consent Form and Free Clarification. After that, the participants were taken to the room containing the eye-tracking equipment and was positioned approximately 70 cm away from a 22-inch monitor with the equipment for recording eye movements. The WMT-2 test was explained to the participant and then they answered the test. At the end of WMT-2, the BRIEF-A, Corsi Blocks Test, Tower of London Test, and WCST were applied. At the end of the procedure, the participant received course credit, a credit necessary for students to graduate, as a contribution to their participation.

Eye-tracking measures

The-average eye-tracking ratio was 95.7% (SD = 4.40). The participant with the lowest tracking ratio had 77.0% of the eyes detect in the task while the participant with the most tracking ratio presenting a tracking ratio of 99.7%. We used the same measures used in study 1. A description of them can be found in Table 4.1.

Data analysis

To test our hypothesis, we conducted several Steiger’s tests to compare the correlations of each predictor with the three cognitive test measures (Tower of London test, Corsi block-tapping test, and WCST perseverative errors). To calculate the Steiger’s test, it is necessary beforehand to calculate the correlation coefficients (r). Accordingly, we calculated the Pearson correlation coefficients between the three hypothesized eye-tracking measures (i.e., Ratio of time spent on the matrix vs answer choices, Number of matrix-answer transitions, and the Average number of visits to a given incorrect answer choice) and the three cognitive tests. Steiger's test statistically compares different correlations coefficients in the same sample by calculating a z-value from the r, evaluating each difference with an asymptotic z-test, and then inferring the p-values. By convention, a significant difference between correlations coefficients is found when the test reveals a z-value greater than 1.96 in two-tailed tests, and therefore a p-value under 0.05 (Steiger, 1980).

We also conducted four LASSO regressions to select from all eye-tracking measures, which, if any, predicted the cognitive test measures and the self-reported EF. To do it, we did data split in our sample: ~80% of the data (57 participants) were used to train the model, and find the best value of the penalty, with a Leave one out Cross-Validation, and ~20% of the data (12 participants) was used to validate the model. We evaluated our model in the test set and calculated the correlation coefficient, R^2, the MAE, and the RMSE.

Results

Descriptives

The descriptive of the variables used in this study are presented in Figure 4.2. In the Tower of London, the participants in our sample presented a mean (SD) score of 32.58 (2.48), ranging from 25 to 36. In the Corsi block-tapping test, the sample's mean was 12.77 (2.94), with participants having scores between 8 and 20. Participants had, on average, 6.99 (6.63) perseverative errors in the WCST. The perseverative errors ranged from 0 to 26. Furthermore, our sample scored 166.59 (27.92) in the BRIEF-A total score. The minimum score was 119 and the maximum score was 240. In the WMT-2, the participants presented a mean of 11.06 (3.61) items answered correctly. The score ranged between 2 to 18.

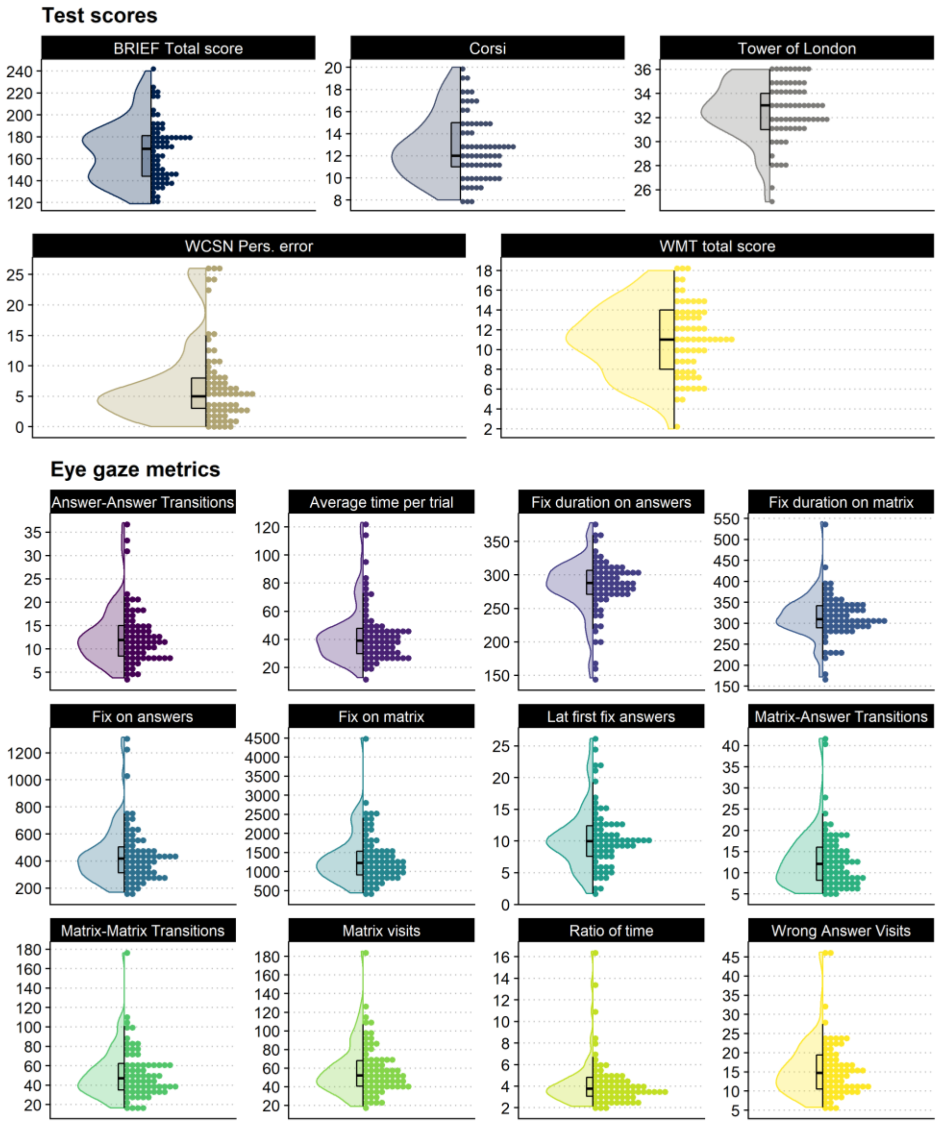

Figure 4.2 – Raincloud plots for the variables used in Study 2.

In the clustering analysis, both strategies found were similar. The first strategy presented a pattern of following each row cell until going to the answers. The second strategy had a similar pattern; however, participants that used this strategy also had a pattern to follow the columns of the matrix. Participants who tended to adopt the second strategy had a higher probability to go to the answer choices from the end of each row, a pattern not found in the first strategy. The similarities and differences of both clusters are shown in Table 4.4. Plots of the transition matrix and of the distribution of different eye-movement measures for each cluster can be found in the supplementary file.

Table 4.4 – Comparison of both non-supervisioned clusters.

Measures	Cluster 1	Cluster 2	Analysis
Direction	Row-wise	Row-wise & Column-wise	Visual inspection (see supplementary file)
Probability to go to the answers when in top and mid row	Low	Low to moderate	Visual inspection (see supplementary file)
Average time spent in each trial	-	-	BF_{10} = 0.07 (±<0.00) ºº
Matrix-matrix transitions	-	-	BF_{10} = 0.40 (±<0.00)
Matrix-answer transitions	-	-	BF_{10} = 0.06 (±<0.00) ºº
Answer-answer transition	Less transitions	More transitions	BF_{10} = 72.79 (±<0.00)**
Latency to the first fix. on an answer choice	-	-	BF_{10} = 0.06 (±<0.00) ºº
Ratio of time	-	-	BF_{10} = 0.08 (±<0.00) ºº
Visits to a given matrix cell	-	-	BF_{10} = 0.32 (±<0.00) º
Visits to a given incorrect answer choice	-	-	BF_{10} = 0.83 (±<0.00)
Fixations on matrix cells	-	-	BF_{10} = 0.08 (±<0.00) ºº
Fixations on answer choices	-	-	BF_{10} = 1.91 (±<0.00)
Average fixation duration for a matrix cell	Shorter fixations	Longer fixations	BF_{10} > 1000 (±<0.00)***
Average fixation duration for an answer choice	Shorter fixations	Longer fixations	BF_{10} > 1000 (±<0.00)***

**Strong evidence for H1, ***Extreme evidence for H1, ºModerate evidence for H0, ººStrong evidence for H0.

Comparing the correlations

To investigate our hypothesis, we conducted a Steiger test between each of the predicting variables that we hypothesized and the three cognitive tests: Tower of London, Corsi tapping-block test, and WSCT perseverative error. To do it, it is necessary to calculate the correlation coefficients beforehand. The correlation coefficients between these measures can be found in Figure 4.3.

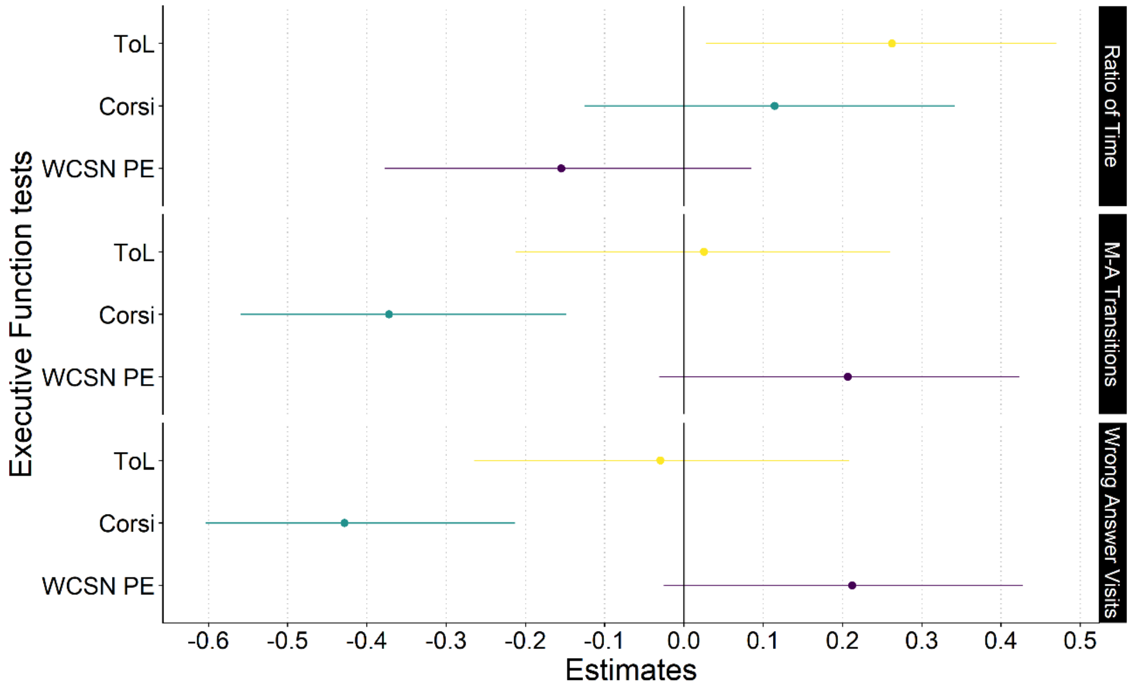

Figure 4.3 – Forest plot presenting the correlation estimates and their 95% confidence interval for the correlation between the select eye gaze metrics and performance on cognitive tasks.

Legend: Ratio of time = Ratio of Time spent on the matrix vs. answer; M-A Transitions = Matrix-Answer Transitions; ToL = Tower of London test; Corsi = Corsi block-tapping test; WCSN PE = WCSN Perseverative Errors.

To analyze the relation of the eye-tracking measures with each executive function test and if there was a statistical difference between them, we performed a Steiger test for each eye-tracking measure. On one hand, the correlations of the Ratio of Time spent on the matrix vs. answer with cognitive tests trended towards significance, $z = 1.74$, $p < 0.08$. On the other hand, the correlations of Matrix-Answer Transitions with cognitive tests presented significant differences, $z = 2.2$, $p < 0.03$. Lastly, the correlations of Wrong Answer Visits with the executive function tests presented no significant difference, $z = 0.32$, $p < 0.75$.

LASSO regression Models

From the four proposed LASSO regression models, three were able to find predictors for the dependent variable. The coefficients of each selected variable and the model measures are displayed in Table 4.5. Latency to the first fixation in answer choices, the ratio of time spent on the matrix vs. answer, and mean fixation duration in the matrix predicted 16% of the variation in the Tower of London test. The Visits in the wrong answer choices were able to predict 17% of the variation in the Corsi test. The model with WCST perseverative errors as the dependent variable did not identify any predictors. Lastly, seven eye-tracking measures were able to predict 36% of the variation in the BRIEF-A total score. It is noteworthy that higher scores in BRIEF-A indicate executive dysfunction, which means that negative coefficients are related to better daily executive functions.

Table 4.5 – Coefficients and performance estimates of the LASSO regression model predicting the cognitive tests performance, and the self-reported EF.

Measures	**TOL**	**Corsi**	**WSCT Perseverative Errors**	**BRIEF-A**
Predictor's coefficients [a]				
Latency to first fixation in answer choices	0.14	-	-	-
Ratio of time	0.02	-	-	-0.25
Mean fixation duration in matrix	-0.13	-	-	-0.18
Visits in wrong answer choices	-	-0.25	-	-
Answer-answer transitions	-	-	-	-0.33
Mean fixation duration in answer choices	-	-	-	0.20
Percent of trials classified as cluster 2 scanpath	-	-	-	0.15
Performance estimates				
Correlation coefficient	0.40	0.41	-	0.60
MAE	0.69	0.59	-	0.58
RMSE	0.92	0.69	-	0.74

[a] Showing the predictors selected by the LASSO model; see full set of eye gaze metrics in Table 4.1.

General discussion

Our study aimed to explore the relationship between cognitive abilities and eye-tracking measures related to strategy use in matrix reasoning tasks. Our results indicated that there is a relationship between several eye-tracking measures with different cognitive abilities.

The Steiger's test was used to test our hypotheses. We hypothesized that the ratio of time spent on the matrix vs. the answer choices would be more related to planning than the other cognitive measures (Hy1). Indeed, we found a trend toward significance in this hypothesis, pointing to the possibility that planning is manifesting through more time spent gazing on the matrix. We also hypothesized that fewer gaze transitions between the matrix and the answer choices would be related to higher WM scores (Hy2). The Steiger's test confirmed this hypothesis. Lastly, we predicted that fewer perseverative errors on WCST would be more related to less revisit in incorrect answer choices. This hypothesis was not confirmed. Based on the Steiger's test, no statistical significance was found pointing to difference between correlations of the cognitive measures.

We also conducted exploratory analysis in order to investigate the relationship of the eye gaze measures with the cognitive tests. To our surprise, the most similar test was the lowest predicted. The results in study 1 showed a relationship, albeit a weak one, between two eye gaze metrics—the number of matrix-answer transitions and the ratio of time spent on the matrix vs. answer choices—and the D.70 score. These variables predicted 6% of the variation in the D.70 test. These results are consistent with previous studies that found a relationship between the matrix-answer transitions and time spent on the matrix with the Gf score (e.g., Hayes et al., 2011, Laurence et al., 2018, Vigneau et al., 2006). However, these studies used the score of the test that the eye gaze was recorded as the fluid reasoning score. However, it is expected that different eye movements in a task would result in a different performance. In our study, we tried to predict a Gf score that was not the same as the one that the eye movements were recorded. We were able to find a relationship between fluid reasoning score and matrix-answer transition with a score of a second reasoning test. It is possible to affirm that the results of Hayes and colleagues (2011), Laurence et al. (2018), and Vigneau and collaborators (2006) were overfitted since they were predicting the outcome inside the same task that the eye movements were recorded. Therefore, our results demonstrate a relationship between the number of matrix-answer transitions and time spent on matrix with fluid reasoning performance, even if the fluid reasoning is measured in a test that was not the one that the eye gaze was recorded, but this relationship is small compared to previous studies.

The cognitive measures and the BRIEF-A had a bigger variance predicted by the eye movements. Higher scores in the Tower of London, a cognitive measure of planning, were related to higher Latency to the first fixation in answer choices, and the ratio of time spent on the matrix vs. answer, and a smaller mean fixation duration in the matrix. It is possible to assume that participants that show better planning abilities are the ones that first try to solve the problem on the matrix and then go to the answer choices, similar to a constructive matching strategy. This is in line with our hypothesis (Hy1). With this in mind, the eye gaze measures selected were all related with the constructive matching: a high latency to the first fixation in answer choices points out that participants were scanning the matrix before going to the answer; a higher ratio of time demonstrates that participants spent more time fixating in the matrix than in the answer choices. Shorter fixations were also related to planning. Hodgson and colleagues (2000) demonstrated that participants that had better performance in the Tower of London test were the ones with less fixation time, so it is possible to assume that good planning abilities are capable of doing fast and efficient fixations.

Further, Higher scores in Corsi block-tapping test were related to fewer visits in wrong answer choices. This result was not expected by us, but it is justifiable. Participants that present higher WM is capable of memorizing the trial and visiting the wrong ones fewer times. This is a result of remembering the matrix patterns and easily identifying wrong answers and remembering the place of the wrong answer and then fixating fewer times there. This result is not in line with our hypotheses (Hy2). We expected higher WM scores to be related to fewer matrix-answer transitions since participants with higher WM would not need to revisit the matrix multiple times in order to remember the patterns. This hypothesis was based on previous work that demonstrated that individuals with higher WM used the constructive matching strategy more times (Gonthier & Roulin, 2020). Conversely, we expected that visits in wrong answers would be related to perseverative errors in the WSCT (Hy3), but our hypothesis proved to be not true. A possible explanation for this is that our sample had a relatively small variation in the WCST perseverative error measure. Maybe, in a more diverse sample that is not based on university students, the variation in the perseverative errors would be higher and a predictor could be found. However, more studies are needed to answer this point.

The BRIEF-A model was even more compelling than the Tower of London and Corsi block-tapping test models. A high variance in the BRIEF-A was explained by the eye-movement measures. Interestingly, this model was the one that selected most variables. This result can reflect the idea that BRIEF-A is measuring daily EF in general, and therefore several eye-tracking measures are related to this broad spectrum of EF because this broad spectrum of EF is influenced by different eye behaviors during the performance. Further, longer fixations were related to a better daily EF. Other studies (Luke et al., 2018, Meghanathan et al., 2015) demonstrated that individuals with high WM tended to do longer fixations in scene and free viewing. Since BRIEF-A accesses different EF, including WM, these long fixation durations may be related to this aspect of the measure.

The selected predictors of the BRIEF-A can also be analyzed through the lens of constructive matching and response elimination strategies (Vigneau et al., 2006). In this sense, spending more time in the matrix and doing fewer transitions are related to constructive matching, which indicates that participants with higher EF are prone to use constructive matching more often. This result is in line with other studies that correlate WM with cognitive strategies and found that individuals with higher WM tend to use constructive matching more often, even in hard items (Gonthier & Roulin, 2020).

One interesting finding is that better daily EF is related to a higher percent of trials classified as cluster 1 scanpath. This scanpath presents less probability to go to the answer choices while they are not in the bottom line of the matrix. This is an interesting result because it is relating the scanpath, accounting for the sequence of fixations, with the self-reported EF. It is the first paper that used Kucharský and colleagues' (2020) scanpath classifying algorithm in a matrix reasoning task and found a relation with cognitive abilities. This demonstrates evidence for validity for this measure.

The results presented in this study have bigger implications for understanding the relationship between eye movements in matrix reasoning tasks and cognitive abilities. It demonstrates that the cognitive strategies in matrix reasoning tasks, measured by eye-tracking, have an average relation with cognitive abilities of planning and WM. It also relates to daily life self-reported EF, which is interesting when looking into implications of research involving these eye movements. If the strategies are related to daily life EF, interventions that teach the participant how to use constructive matching strategy may hone daily life EF? Or the gains related to learning constructive matching will transfer to the cognitive abilities? New studies seeking to answer these questions are needed.

This study was not carried without limitations. First, we had two different samples with different cognitive measures instead of one with all the cognitive measures. Second, both of our samples consisted of university students, which is a concern for the generalization of the results. Future studies should focus on diverse samples.

To conclude, we aimed to explore the relationship between cognitive abilities and eye-tracking measures related to strategy use in matrix reasoning tasks. To our surprise, the Gf test was the one less predicted by the eye-tracking measures. The self-reported EF, WM, and Planning were respectively the ones that the eye-tracking measures predicted higher variance. No eye-tracking measure was a predictor of WCST perseverative errors. With these results in mind, it is possible to assume that the cognitive visual strategies used in the matrix reasoning task are influenced by cognitive abilities such as planning, WM, fluid reasoning, and EF.

Supplementary Information

S1. Changes after preregistering the analysis.

In the last paragraph of Question 5 in the AsPredicted, we mentioned that we would do exploratory analysis to predict the performance in the different cognitive tests. However, we decided to use a Least Absolute Shrinkage and Selection Operator (LASSO) regularization

(L1) instead of the Elastic-Net regularization (L1+L2). Our change occurred because we were interested in selecting different eye movement predictors and the L2 regularization would not help with it. Therefore, the penalization without exclusion of any of the independent variables by the L2 regularization would not be useful to us. Consequently, we decided to only use the L1 regularization.

Study 1: Cluster Analysis

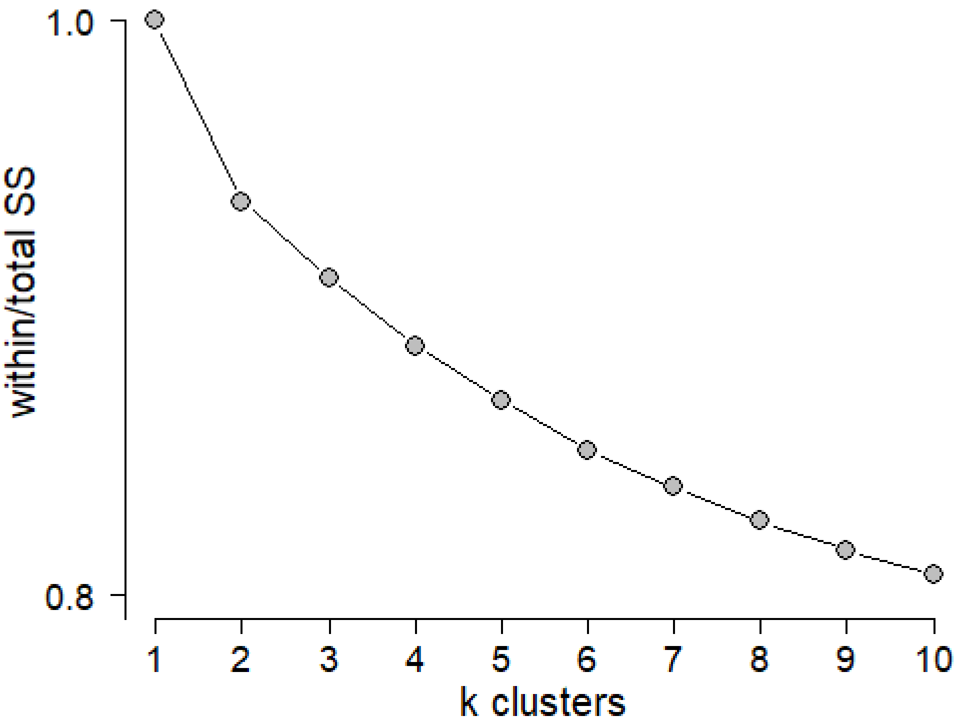

Figure S1. The scree plot of the clusterization model. In this plot, it is possible to see that there is no evident “elbow” on the plot. Therefore, since the literature is based on 2 strategies (e.g., constructive matching and response elimination), we decided to use k = 2.

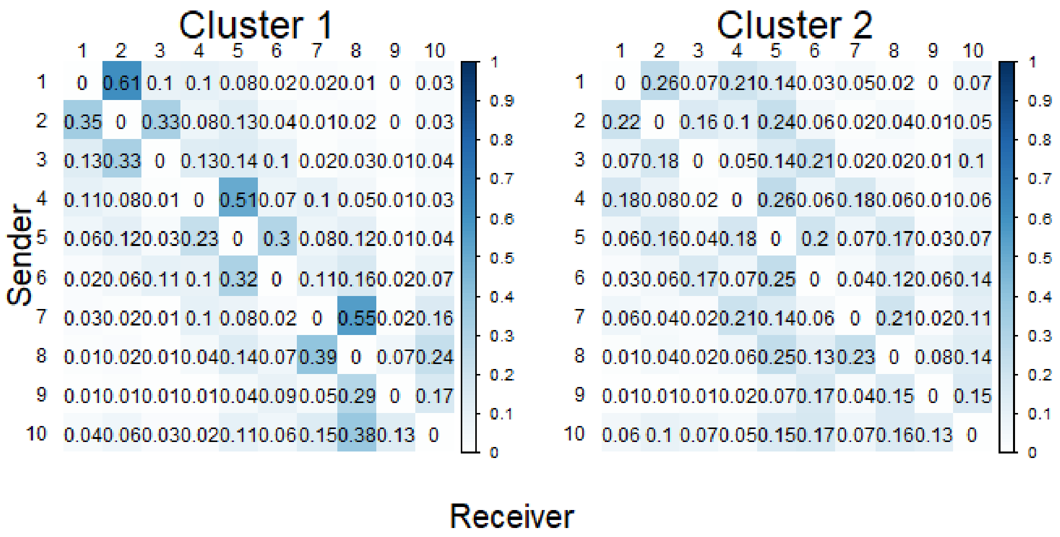

Figure S2. Average transition matrix plot of the clusters. Each row is the sender, and each column is the receiver. Higher values (and blueish colors) imply a higher probability that the fixation that occurred in that row will go to the cell indicate by the column. Cluster 1 tends fixations in each line of the matrix task, while cluster 2 has a tendency of line and columns. Cluster 2 also has a higher chance of going to the answers (AOI 10) while under cells 1 to 6.

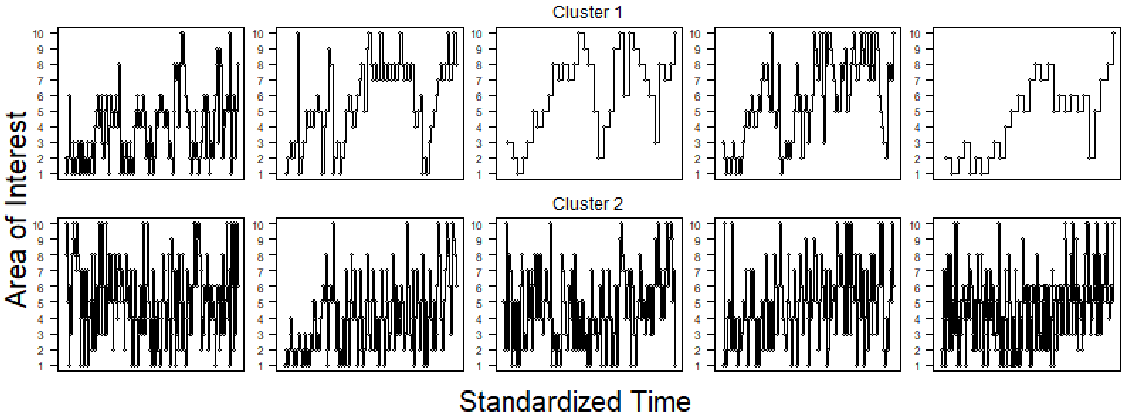

Figure S3. Representatives of the eye movements categorized in each cluster. Each plot is the sequence of fixations in the AOIs while in one trial of the matrix reasoning task.

Study 2: Cluster Analysis

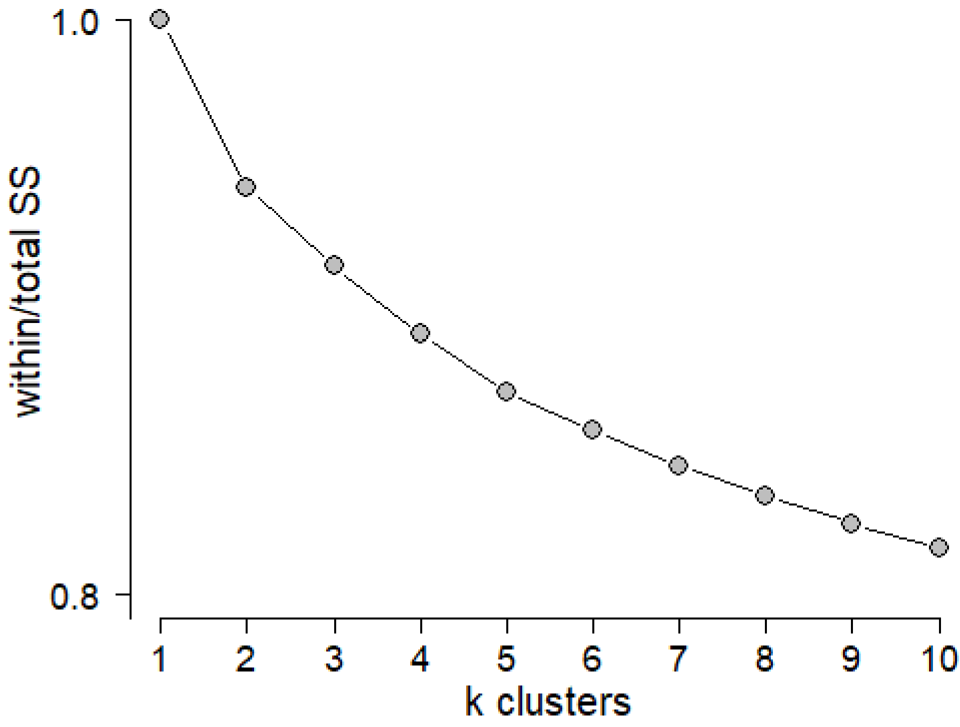

Figure S4. The scree plot of the clusterization model. Similar to the scree plot of study 1 (Figure S1), it is not possible to see a clear “elbow”. Therefore, we followed the same concept of study 1 and used k = 2 based on the literature.

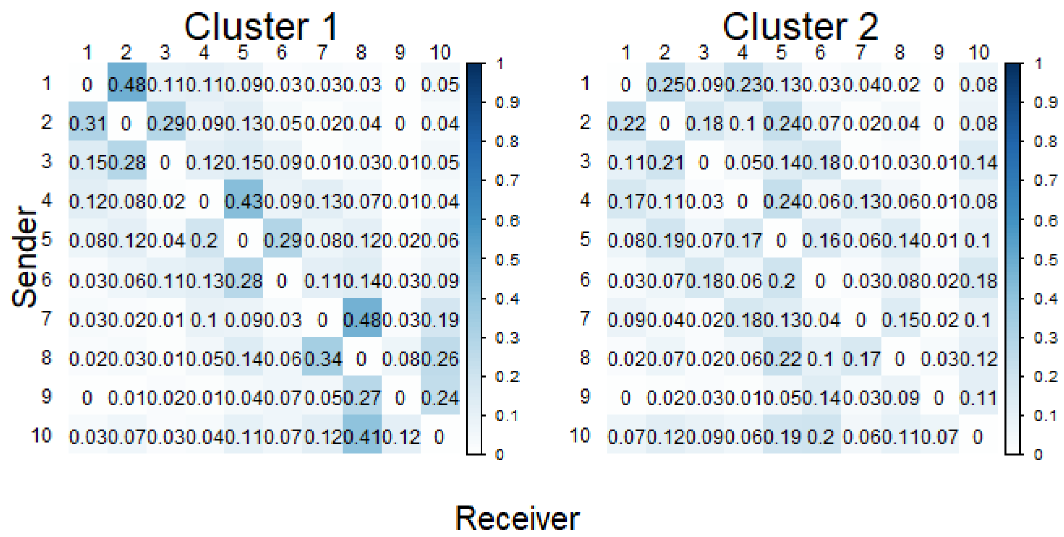

Figure S5. Average transition matrix plot of the clusters. Each row is the sender, and each column is the receiver. Higher values (and blueish colors) imply a higher probability that the fixation that occurred in that row will go to the cell indicate by the column. Similar to the data found in study 1, cluster 1 has a tendency of fixations in each line of the matrix task, while cluster 2 has a tendency of line and columns. Cluster 2 also has a higher chance of going to the answers (AOI 10) while under cells 1 to 6.

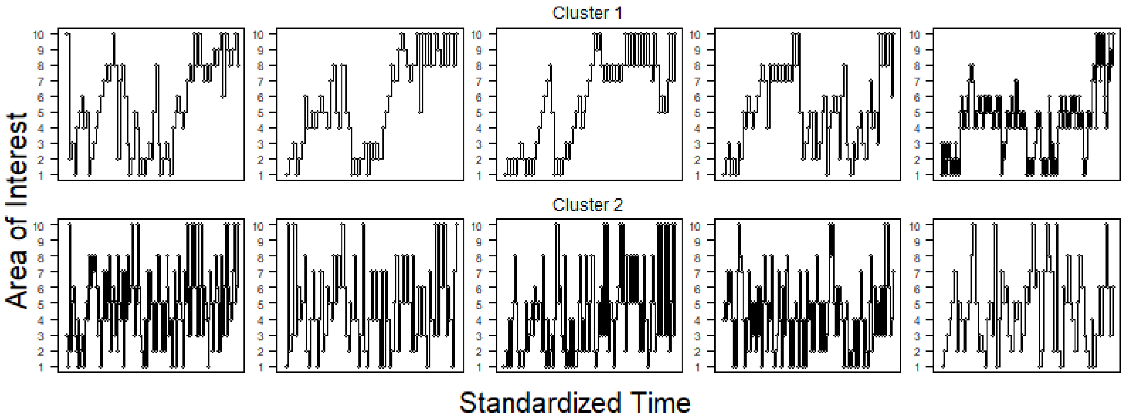

Figure S6. Representatives of the eye movements categorized in each cluster. Each plot is the sequence of fixations in the AOIs while in one trial of the matrix reasoning task.

CHAPTER 5: COGNITIVE VISUAL STRATEGIES IN THE CLOZE TEST: RELATIONSHIP WITH WORKING MEMORY AND PERFORMANCE [3]

Introduction

Thanks to technological advances and the increased processing power of computers, it is possible to use eye movement tracking techniques to identify areas that participants are looking at during a computer task (Płużyczka, 2018). With ever-increasing resolution in eye tracking technologies and the use of advanced data analysis techniques, it is currently possible to extract data from a participant's eye movements during a task and then process this data in order to achieve numerical measurements that can be used to infer the pattern of cognitive strategies employed (e.g., Kucharský et al., 2020; McCray & Brunfaut, 2018). Until the last decade of the last century, quantitative measures of cognitive-visual strategies were extremely difficult to obtain, with qualitative work on the classification of visual tracing patterns being more common (e.g., Snow, 1980).

Studies on cognitive strategies drawn from the analysis of eye movements are a necessary effort to understand how cognitive processing in different tests and situations works, but there are still only a few studies that fully address these issues. In reading comprehension tasks studied through the use of cloze-type tests, the first article which worked directly with eye movements in a systematic way was only published in 2018 (McCray & Brunfaut, 2018), although other studies have previously hypothesized about the topic (Gao & Gu, 2008; Yamashida, 2003). This indicates the novel nature of the topic.

In reading tests, a very common task used to analyze cognitive strategies is the cloze test. The cloze test, also known as gap-filling test, is made out of a text with some words omitted, the participant has to fill these gaps with the words they deem to fit better. In the fixed-ratio cloze test, the word that would be in the 5th, 7th, or 10th position in the text is systematically omitted (Oller & Jonz, 1994). The banked cloze has the answers that were omitted from the text in a word bank, the participants need to select words from this bank to reconstruct the sentences (see Figure 5.1; Anderson & Csereszny és, 2003). The cloze test is a test that is often used for second-language learning (e.g., Tremblay, 2011) and it can be used to access a language proficiency for a native language and for a second language (Alderson, 2000). Despite some

[3] The text in this chapter is based on a paper coauthored by the author, Stella A. Bassetto, Natalia P. Bertolino, Mayara S. C. V. O. Barros, and Elizeu Coutinho de Macedo.

critics, cloze tests are a very interesting way to measure text difficulty. Since all the items are similarly structured, what changes between each item is only the text difficulty. Furthermore, as the cloze tests solely and objectively ask for the participant to fill out the gaps in the text, it avoids researcher bias commonly associated with questions about the text (Klejin et al., 2019). The cloze test is a good example of a task that requires different steps (e.g., fill the gaps) in order to complete it, but these steps give researchers little information regarding cognitive strategies. However, cognitive strategies in the cloze test can be analyzed through eye-tracking.

Figure 5.1 – An example of a banked cloze test item in Portuguese with the 5th word omitted. The area of interest of the text and the word bank are highlighted.

Although few studies have addressed the cognitive strategies employed during the cloze test, two types of strategies have been described in the literature: the first strategy refers to when a person searches information directly in the gaps of the test and surrounding words, in a more local pattern, in order to try to quickly fill in the missing parts with the words available in the word bank; the second strategy refers to when a person tries to read the text as a whole and then look for words in the word bank that fit the text in a sentential, or global, pattern (Gao & Gu, 2008; Yamashida, 2003). The first strategy has been addressed as a local strategy while the second one has been addressed as a sentential or global strategy (McCray & Brunfaut, 2018).

In practical terms, it is possible to characterize these strategies based on eye movements. For example, in the global strategy, the participant is expected to spend a greater amount of time in the text than in the word bank, when compared to a local strategy. Logically, in the local strategy, the participant is expected to spend a greater proportion of time on the word bank, compared to the global strategy. Finally, it is expected that people who present a global strategy have fewer toggles between the text and the word bank. This happens because, as the participant will have a greater understanding of the sentence, he will search for the ideal word in the word bank, whereas in the local strategy, multiple comparisons will be necessary for the participant to be able to answer (McCray & Brunfaut, 2018).

These strategies can also be found through scanpaths. In this sense, instead of collecting data from certain eye behaviors, such as time spent in an area of interest (AOI) or toggles between AOIs, it is possible to analyze each fixation in a sequence, taking time into consideration. Therefore, it is possible to understand the sequence of fixations and the path that each person took to analyze the task. Nowadays, it is possible to cluster these scanpaths in groups based on their similarity. To do this, it is necessary to use non-supervised classification algorithms. These algorithms will find the scanpaths that are more similar to each other and define them as a cluster. One of the benefits of this type of approach is that the algorithm will classify the scanpaths by itself, so the researcher only needs to interpret them. Another benefit is that this type of classification will consider time, thus considering the sequence of fixations that a participant made, which is an important aspect of eye-tracking data. A good example of this approach is the method created by Kucharský and colleagues (2020). In this method, eye-gaze data will be converted into a transition matrix of each AOI. With this transition matrix, it is possible to find the most similar patterns and aggregate them into a cluster using k-means clustering. This method will produce an average transition matrix based on the hidden Markov model. With this, the researcher can visualize this transition matrix and understand which underlying pattern the algorithm captured and, then, interpret each cluster.

Evidently, the adoption of a strategy in the cloze test is related to reading abilities. Brown (2003) hypothesized that given the reading ability of a person, participants will execute different strategies. On one hand, lower proficiency readers would do a more local, lower processing, strategy because it is hard for them to handle complex language information. On the other hand, higher proficiency readers would do a global, higher processing, strategy since they can understand complex textual information.

This hypothesis is in line with Khalifa and Weir's (2009) cognitive model of reading. This model takes into consideration a bottom-up perspective that begins with visual input. In the lower levels, simple and automatic processes take place, such as word recognition and lexical access, while in the higher levels, strategic and complex processes are executed. In higher levels, there are processes such as syntactic parsing, establishing propositional meaning, inferencing, building a mental model, and creating a text-level representation. Since it is a hierarchical model, the lower-level process must be accurate in order to achieve efficient reading comprehension. Consequently, problems in word recognition would affect higher-order processes and generate difficulties in reading comprehension. Khalifa and Weir (2009) also suggest a top-down process that uses context information to help text comprehension. When a person has problems in the decoding of a text, they are more likely to use this top-down process to help them, making them less available to capture the global meaning of a text. Therefore, it is possible to hypothesize that a person with lower proficiency can engage in a local visual cognitive strategy while participants with higher proficiency engage in a global visual cognitive strategy. This was tested by Yamashida (2003) with a small sample of 12 Japanese readers in a cloze test experiment. The author found that less proficient readers presented a local emphasis while more proficient readers presented a global emphasis.

Furthermore, Khalifa and Weir (2009) also mentioned that more attentional capacity available in a person working memory (WM) would convert to better efficient reading comprehension. Therefore, individual differences in WM may influence a person's choice of strategy. The WM has several components, including verbal short-term WM, a system characterized by the quick storage of verbal information, and visuospatial short-term WM, defined as the system that temporarily stores spatial information (Baddeley, 2012). Both verbal WM and visuospatial WM may be important for the cloze test: the first, due to its relationship with quick verbal information, a relevant aspect in the cloze task; while the second is related to localization of different parts in the test, another relevant ability in the test. An example of the WM influence on the cloze test is that a person who has a larger working memory capacity (WMC) should be able to store the text and the word bank information, doing fewer toggles and fewer text-word bank toggles. McCray and Brunfaut (2018) commented on this hypothesis but did not test it as they did not have a WM measure. Thus, to the best of our knowledge, no study has tested the hypothesis that WM is related to strategies adopted in the cloze test.

Few studies have focused on cognitive strategies applied during the cloze test. One of them was the Yamashida (2003) study, mentioned above. Gao and Gu (2008) also studied strategies applied on cloze test. They had a sample of 18 participants (divided into three groups – low, medium, or high reading scores – with 6 participants each). In their study, participants used mostly a local strategy instead of a global strategy, which is the opposite of what was found by Yamashida (2003). This difference may be due to Gao and Gu's (2008) study having a word bank while Yamashida's (2003) study did not had a word bank. Nevertheless, both studies were carried without using eye-tracking. McCray and Brunfaut (2018), to the best of our knowledge, was the first study to analyze the cloze test performance while recording eye movements. In this study, the authors used a sample of 28 participants to investigate the processing of test-takers during a banked cloze test. They found that lower-scoring participants used a local eye-movement strategy while high-scoring participants had a global eye-movement strategy.

Analyzing the literature regarding cognitive strategies applied to cloze test, some gaps can be identified. First of all, the studies had a small sample, with no study having more than 28 participants. Furthermore, only one study used eye-tracking. However, even in this study, the measures did not account for the scanpath and the richness of the sequence of fixations. Instead, the eye-tracking measures were a summarize of toggles, time on the text and word bank, and fixations. Lastly, studies mentioned and hypothesized the relationship between strategies, performance, and WM. But, to the best of our knowledge, no study had tested it.

With these gaps in mind, the aim of the present study was two-fold: first, we aimed to identify the cognitive visual strategies in the cloze test using a non-supervised algorithm that accounts for the scanpath instead of the summarized events; second, we aimed to analyze the relationship of these strategies with the working memory and the performance in the cloze test.

These objectives are justified since the cloze test is a very common task in language learning, especially for English, therefore a study understanding how test-takers process this task is relevant. Furthermore, different insights can be understood from how humans read and solve complex reading tasks when we analyze their eye movements (e.g., Kucharský et al., 2020; McCray & Brunfaut, 2018). Thus, a study investigating the scanpath of participants in this task can benefit future researchers in understanding how we solve complex reading tasks and how we process reading a text.

Methods

This study was preregistered in the AsPredicted platform under registration number 66445 before data access and it can be visualized at https://aspredicted.org/blind.php?x=az5nn3. All deviations from the initial plan are described in Supplementary File 1. The data, codes, and analysis are also openly available at https://osf.io/sgxkh/?view_only=918c1c46edf14b0482cb8f51793cbd32.

Participants

For this study, data was collected with a sample consisting of 51 Brazilians, they were Portuguese speakers, young adults, and university students of both sexes (23 women, 45.09%). The average age of the participants was 23.27 years (SD = 2.92), and their ages ranged between 18 and 31 years. Three participants chose not to disclose their age. The majority of these students were studying law (38; 74.5%), followed by psychology (10; 19.6%), publicity and advertising (2; 3.9%), and only one student studied pharmacy (2.0%). The data was collected between February and April of 2019.

For the inclusion criteria, it was considered whether the participant had typical or lens corrected vision. As an exclusion criterion, it was considered whether the participant had ever been diagnosed with a neurological or psychiatric disorder and whether they presented a score above 12 on the Adult Dyslexia Checklist or a WM scores over 3 standard deviations.

Instruments

Adult Dyslexia Checklist (ADC)

The Adult Dyslexia Checklist questionnaire (Vinegard, 1994) consists of 20 items that evaluate different aspects of dyslexia. Each item consists of a sentence (e.g., “Reading a map or finding your way in a strange place does it seem confusing?”), to which participant must answer either “yes” or “no”. All affirmative responses count as a point for the total score of this questionnaire. The version of ADC used in this study had been previously translated into Portuguese (Rosa, 2016). This instrument was used to assess indicators of dyslexia among the participants.

Cloze Test

The Cloze Test is a text reading test in which the participant is asked to “fill the gaps” presented throughout the paragraphs with words from a word bank. This test consists of 11 texts that increase in the degree of difficulty and complexity according to the progression of the items. The test was in Portuguese and the complexity of the task - measured by Coh-Metrix-Port 2.0 - increases in terms of text size, frequency, and length of words (Penteado, 2019). Gaps were always inserted on the fifth word, which is replaced by a highlighted open space that is always the same size (see Figure 5.1). The first three texts had 25 words and a total of 5 gaps. The fourth, fifth, and sixth texts had 50 words and 10 gaps. The seventh, eighth, and ninth texts had 75 words and 15 gaps. Finally, the tenth and eleventh texts had 100 words and 20 gaps. The score is calculated by the number of gaps that the participant missed. Thus, the maximum test score is 130 points, one point per gap. Since the number of correct gaps filled tends to have low variance and present ceiling effect because our participants were proficient in Portuguese, we also calculated the inverse efficiency score (IES), which is defined by the time in each item divided by the percentage of correct gaps filled in the item. Low scores indicate scores of higher efficiency while higher scores indicate the opposite (Bruyer & Brysbaert, 2011). To present the cloze test, we used the GoConqr website tool (available at https://www.goconqr.com/).

Corsi block-tapping test

The Corsi block-tapping test is used as a measure of visuospatial working memory. It consists of a 25 cm by 30 cm wooden board that serves as a base for nine small wooden blocks (3 x 3 x 3 cm) which are permanently distributed on top of the board. Each block has a number - ranging from 1 to 9 - on the side of the cube facing the researcher, making the participant unable to see these numbers. The test is split into two parts: during the first half, the researcher will touch the blocks in a sequence and the participant must repeat the same movements of that sequence; during the second half, the evaluator touches the blocks in a sequence and the participant must repeat the sequence in reverse. The total number of sequences is 14, and the difficulty of the tasks increases every two sequences, with an additional touch being added to the sequence. The test ends when the participant misses two sequences with the same number of touches. The total score of the test varies between 0 and 28. The Corsi block test presents favorable evidence of its applicability in a Brazilian context (Santos et al., 2005).

Digit span test

The digit span test assesses the storage capacity of a person's verbal working memory and is included in WAIS-III and WISC-IV. The test is divided into two parts: digits in the forward order and digits in the reverse order. In both parts, the participant is told a sequence of numbers by the researcher and must repeat it after the evaluator finishes. In the first part, the sequence of numbers is said, and the participant repeats in direct order. If the participant responds correctly, the test continues. For every two sequences of numbers, the sequence is increased. In the second part, the participant must repeat the sequence in reverse. In the first and second half, the test is interrupted when the participant misses two sequences with the same number of digits. The test score ranges from 0 to 30 points, according to the number of sequences that were repeated correctly, with 16 sequences in the direct order and 14 in the reverse order. The test showed evidence that it can be used in the Brazilian context (Figueiredo & Nascimento, 2007).

Eye-tracking device

To record the eye movement data, an eye tracking device by SensoMotoric Instruments RED500 was used. The data were recorded at a frequency of 60 Hz per second. The iView^{TM} software (v. 3.7, SensoMotoric Instruments, Inc.) was used to calibrate the eye-tracking equipment and to record the data. Standard 9-point calibration was used. The Experiment Center^{TM} software (v. 3.7, SensoMotoric Instruments, Inc.) was used to create the experiment and present the stimuli to participants in the GoConqr website, while the BeGaze^{TM} software (v. 3.7, SensoMotoric Instruments, Inc.) was used to process the data and extract it. The fixation and ocular data algorithm used was the standard BeGaze™ algorithm with minimum fixations of 60 ms.

Procedure

The project was submitted to the Ethics and Research Committee of the University and approved in December of 2018 (CAAE 98599618.2.0000.0084). Subsequently, contact was made with the participants through online pages associated with the university. The collaborators were invited to the laboratory and informed about the research objectives. In case of interest and consensus, participants gave written informed consent, and all the procedures followed the Declaration of Helsinki.

The first instrument applied was the ADC questionnaire. If the score was outside the exclusion criteria, the procedure was continued. Then, the participant took the Corsi blocks test and the Digits test. After the completion of these two tests, the collaborator was taken to the experiment room that contained the eye-tracking. The participant was positioned approximately 70 cm away from the 22-inch monitor that had the eye-tracking equipment attached. It was then explained how the Cloze test worked, afterwards the participant could ask questions. Then, the calibration procedure was conducted. After the calibration was completed, the participant started the Cloze test. At the end of the procedure, the participant received extra credit hours in return for their collaboration.

Eye-tracking measures

We marked AOIs on the items of the test. We created two AOIs: an area for the text (called AOI-Text) and an area for the word bank (called AOI-Bank). Therefore, we implemented the method from Kucharský and colleagues (2020) in the eye-tracking data from the participants to identify the clusters of eye movement strategies between both AOIs. To do this, we adapted Kucharský and colleagues’ (2020) codes from Progressive Matrices that are openly available at https://osf.io/wvzs9/. We used the algorithm accounting for repeated fixations. We opted to use two clusters, following previous literature (Gao & Gu, 2008; McCray & Brunfaut, 2018; Yamashida, 2003). The Scree Plot (Supplementary File 1) indicated that two clusters were the most adequate number of clusters. We also calculated the number of fixations, the number of toggles between text and word bank, the rate of toggling, and the time ratio between the text vs word bank. The number of fixations corresponds to how many fixations a participant had in the trial. The number of toggles between text and word bank can be defined as the number of times that a participant gazed from the text to the word bank and vice versa. The rate of toggling is the number of toggling equalized by the time that the participant spent on the trial, a higher rate of toggling means that the participant made more toggles per second. The time ratio on the text vs word bank is calculated using the time a participant spent fixating on the text and then dividing it by the time a participant spent fixating on the word bank, thus indicating how many times the participant spent fixating on the text compared to the word bank.

Data analysis

First, we looked for outliers in the WM measures with 3 or more standard deviations from the mean. We used Shapiro-Wilk normality tests to access the normality of the WM measures. We conducted several elastic-net regressions. We used both WM test scores to predict the number of toggles between the AOIs, the time ratio on the text vs word bank, and the strategy cluster in each item. A model to predict the correct answer was also conducted and is available in Supplementary File 1. Both WM test scores and the strategy cluster to predict the efficiency on each item of the cloze test were also used. In this last model, we opted to use a ridge regression instead of an elastic-net, since we wanted to observe the relation of each parameter with the outcome. Lastly, the correlations between the WM test scores with the toggles between the AOIs and the time ratio on the text vs word bank were calculated. In all models, we conducted cross-validation to identify the better fitting parameters and then tested the model with these parameters in 20% of the data (test data) that was held out of the cross-validation. The cross-validation was conducted using the Leave-One-Out method. To analyze if there were significant differences between each correlation, we used the Steiger's Test. All analysis and data treatment were conducted in R (R Core Team, 2021). We also used the *glmnet* package (Friedman et al., 2010) to conduct the elastic net regressions, *ridge* (Cule et al., 2021) to conduct ridge regressions, and the *psych* package (Revelle, 2020) to conduct the Steiger's Test. All data, codes, analysis, and results are available at https://osf.io/sgxkh/?view_only=918c1c46edf14b0482cb8f51793cbd32

Results

We found no outliers in the WM scores; therefore, no participant was excluded. The working memory measures presented a normal distribution (Digit span test $p = 0.086$, Corsi block-tapping test $p = 0.163$). Participants presented a mean score of 14.92 ($SD = 2.99$) for the digit span test and a mean score of 13.06 ($SD = 2.58$) for the Corsi block-tapping test. Regarding the ADC, participants presented a mean of 4.24 points ($SD = 2.55$), and their scores ranged from 0 to 12.

Regarding the Cloze gap-filling test, participants, on average, answered correctly 10 (SD = 1.01) of 11 items. They took about 15 and half minutes to complete the test (*M* = 940.94 seconds, SD = 238.20). Per item, participants had a mean of 210.69 fixations (SD = 169.66) and 46.41 toggles (SD = 33.02) between the AOIs. Their efficiency per item was of 88.221 seconds (SD = 67.736).

Strategies

Utilizing the Kucharský and colleagues' (2020) algorithm, we were able to identify two clusters of eye movement strategies. The first strategy indicates that participants will do a sequence of fixations on the text, then go to the word bank and do few toggles. The second strategy is related to participants that do multiple comparisons between both AOIs. The transition matrix and representatives of each strategy can be found in Figure 5.2.

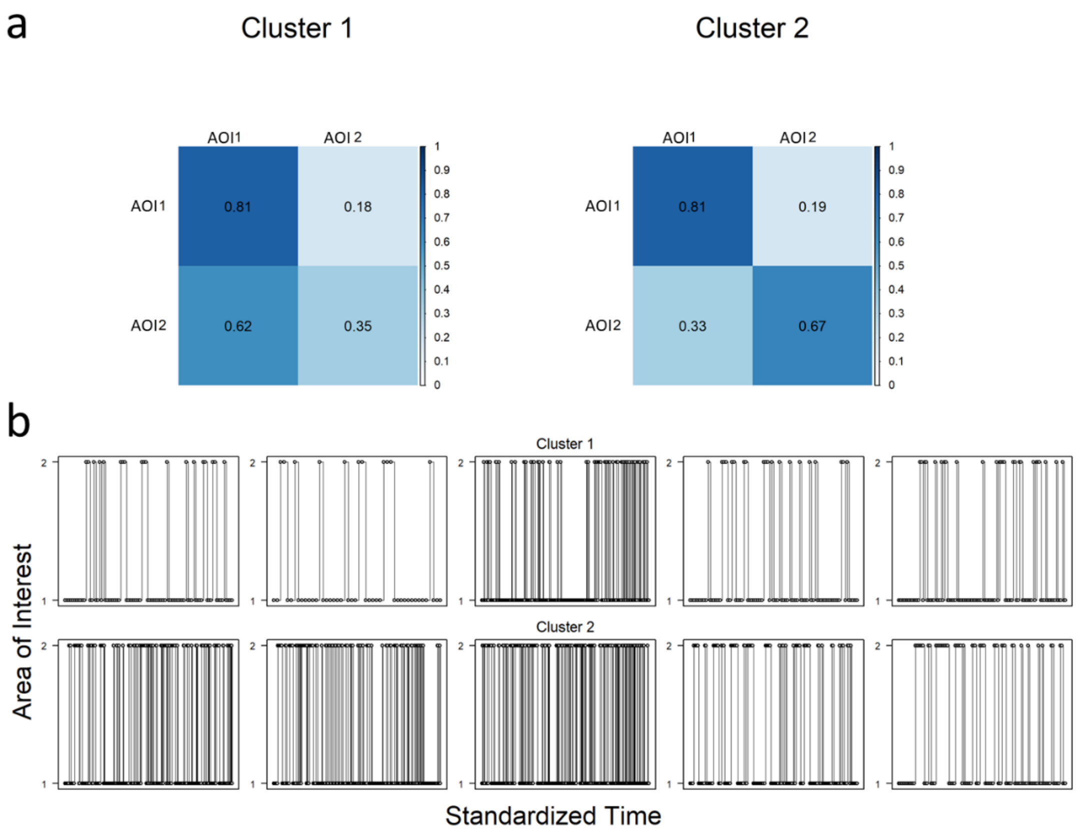

Figure 5.2 – The results from the clusterization of eye-movements. The first figure (a) presents the transition matrices of both clusters, the value inside each box represents the probability of a participant that fixated in the AOI of the line to go to the column AOI. The second figure (b) presents five real examples of eye gazes that were categorized as the first strategy and five eye gazes that were categorized as the second strategy to exemplify each scanpath. The AOI 1 is the AOI-Text and the AOI 2 is the AOI-Bank.

These strategies also present differences in the number of fixations, toggles, rate of toggling, and time ratio on the text vs word bank. Figure 5.3 presents these differences. Cluster 1 presents a low number of fixations and toggles per item, while Cluster 2 has a bigger variability. In Cluster 2, participants can have a higher number of fixation and toggle than in Cluster 1. When equalizing the number of toggles per time, we can see that there is a greater variability of the rate of toggling in Cluster 1. Regarding the time ratio, Cluster 1 has more participants spending more time in the text than participants in Cluster 2.

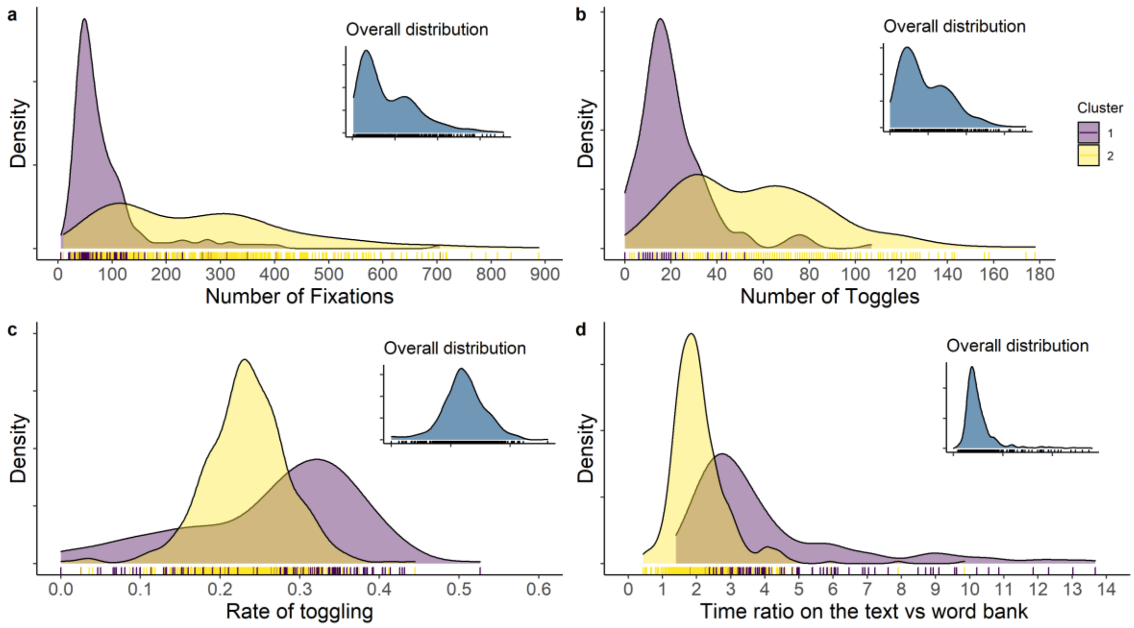

Figure 5.3 – The distribution of the (a) number of fixations, (b) the number of toggles, (c) rate of toggling, and (d) time ratio on the text vs word bank for both clusters.

With 51 participants and 11 items, we were able to analyze 561 scanpaths. From these 561 scanpaths, 162 were classified as Cluster 1, and 399 were categorized as Cluster 2. Their count in each item of the cloze test can be seen in Figure 5.4. It is noticeable the drop of cluster 1 scanpaths between items 3 and 4, 6 and 7, and 9 and 10. All of these were points where the complexity of the text increased. The strategy used by each participant in each item is presented in Supplementary File.

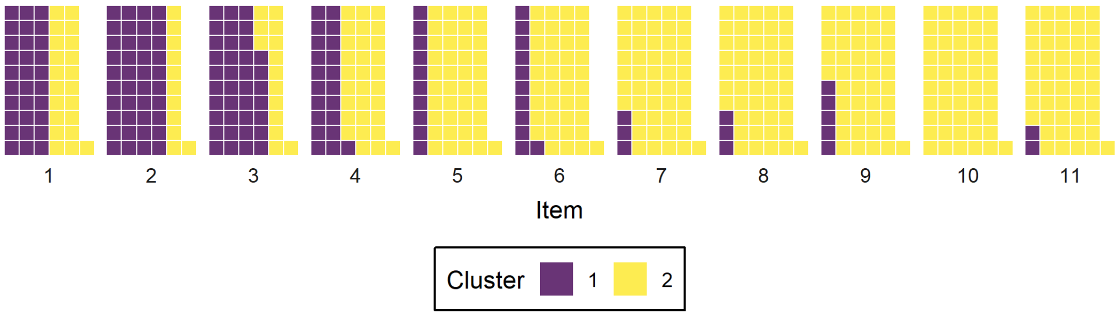

Figure 5.4 – The count of both clusters per item of the cloze test. Each item contains scanpaths of the 51 participants, and each column of the plot has 10 classified scanpaths.

Predicting eye movement measures with WM scores

To understand the relationship between visuospatial working memory and working memory span with the eye-movement measures, we created three regression models. The coefficients and measures of the models can be found in Table 5.1. Additionally, we analyzed the correlations between both WM tests and both eye-tracking measures. Steiger's test was also carried out to compare the correlations of both WM tests with each eye movement measure. Regarding the number of toggles, a marginal significance was found ($t = 1.87$, $p = 0.068$), but no significant difference was found for the time-ratio measure ($t = 0.39$, $p = 0.700$).

Table 5.1 – Coefficients and measures of the models predicting the number of toggles and the time ratio on the text vs word bank.

		Number of toggles	Time ratio on the text vs word bank
Predictor's coefficients	Intercept	530.805	2.576
	Digit Span	-0.861	0.001
	Corsi block-tapping	-	-0.010
Model measures	Alpha	1	1
	Lambda	15.80	0.17
	$R^2_{training}$	0.61	0.54
	MAE	114.64	0.99
Test measure	R^2_{test}	0.30	0.18
CV measures	R^2_{CV}	0.016	0.003

We also used both WM measures and items to predict the Cluster that each participant used in this item. Since the cluster can only be the first or the second, we used an elastic-net logistic classification algorithm. In the training data, we found a cross-validated alpha of 0.8 and a lambda of 0.02. The accuracy of the model was 82.44% (IC95% = 78.61 – 85.85), the sensitivity was 0.70 and the specificity was 0.87. Regarding the predictor's coefficients, the Digit span was not selected by the algorithm, while item (B = 0.472) and Corsi block-tapping test (B = -0.001) were selected. When applying this model to the test data, the accuracy of the model is 80.18% (95% IC 71.54-87.14). The model had a sensitivity of 0.69 and a specificity of 0.84. The ROC curve can be seen in Figure 5.5.

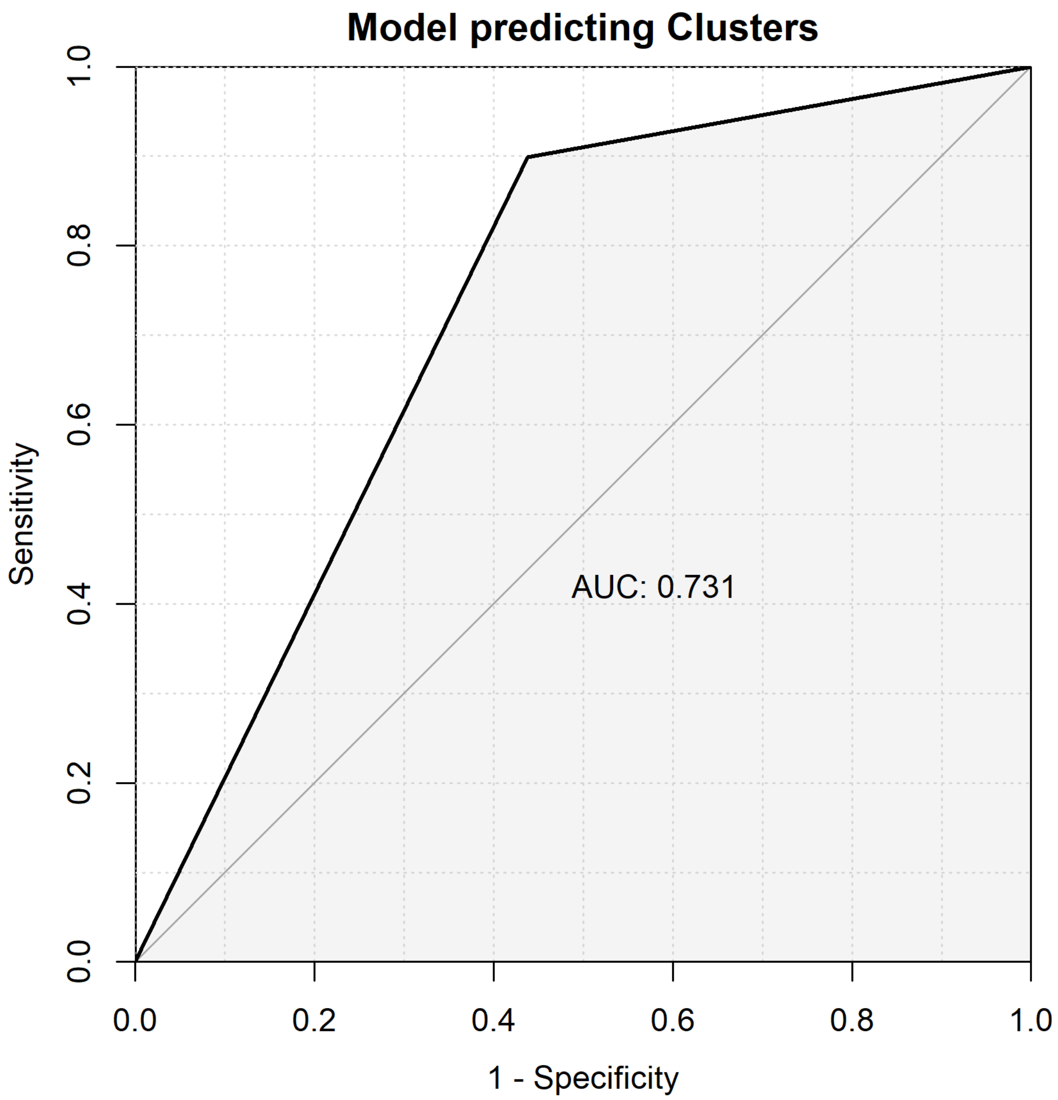

Figure 5.5 – The ROC curve of the model predicting which cluster was used in each item. The Area Under the Curve (AUC) is presented in the plot.

Predicting the efficiency on Cloze test based on strategies and WM scores

A last model was created to verify if the WM scores and the strategy applied could predict the efficiency on each item of the Cloze test. The coefficients and measures of the model are presented in Table 5.2.

Table 5.2 – Coefficients and measures of the model predicting the inverse efficiency score (IES) on each Cloze test's items.

		IES on Cloze test	
		Model	**Scaled Model**
Predictor's coefficients	Intercept	43.030	-
	Item	13.967	185.557***
	Cluster 1	-4.379	-89.387***
	Digit Span	-2.347	-33.049**
	Corsi block-tapping	-0.107	-13.130
Model measures	Lambda	4.13	
	$R^2_{training}$	0.48	
	MAE	32.71	
Test measure	R^2_{test}	0.55	
CV measures	R^2_{CV}	0.249	

Discussion

To recapitulate, this work aimed to identify cognitive visual strategies applied in the cloze test using a non-supervised algorithm that accounts for the scanpath, and to analyze the relationship between these strategies with WM and efficiency in the cloze test. Our study demonstrates that the use of non-supervised clustering algorithms for eye gaze data can provide insights into cognitive processes in a gap-fill task.

First, we want to highlight that our study had the biggest sample in studies that investigate strategies in a cloze test. Other studies had 12, 18, and 28 participants. Our study had 51 participants, almost double of the second study with more participants.

Regarding the strategies, we found two clusters. The first presents a pattern where the participant has multiples fixations on the text, then goes to the alternatives few times and then returns to the text. The second pattern indicates multiple comparisons between text and word bank. Also, in the second cluster, participants tend to do more repeated fixations in the word bank than in the first strategy. A big difference between both clusters is concerning the number of fixations, toggles, and time spent on the text. Participants that presented the first strategy tended to do fewer fixations and fewer toggles than the participants who employed the second strategy. Furthermore, participants were using both strategies at the beginning of the test, but as it progressed and more complex items were presented, participants shifted almost exclusively to the strategy of Cluster 2. It is noteworthy that the number of Cluster 1 decreases as the items change in complexity. For example, between items 3 and 4, items 6 to 7, and items 9 to 10 there is a steep decline.

These clusters were very similar to strategies mentioned in previous literature (Gao & Gu, 2008; McCray & Brunfaut, 2018; Yamashida, 2003). The first cluster can be defined as a strategy that involves a more global understanding of the text, connected with higher processes; and this was evident in the high number of fixations in the text at the start, and then searching for answers in the word bank. The second cluster can be defined as a local strategy, related to lower processes, where participants will do several comparisons to try to find the answers locally instead of accessing the global understanding of the sentence. These patterns are reflected by the higher number of toggles compared to the first cluster.

These strategies present some associations with WM. This is an expected feature since participants with higher WM should be able to memorize the text and the words on the word bank and then do fewer toggles, for example. This possibility was discussed in the literature (McCray & Brunfaut, 2018) but, to the best of our knowledge, no study addressed this hypothesis until now. Our data suggest that a higher WM span is indeed correlated to fewer toggles while the visuospatial WM is not a predictor. A tendency for a significant difference was found between both of these measures with the number of toggles, indicating that the correlations between the digit span test and the Corsi block-tapping test with the number of toggles were different.

The same principle applies to the time ratio between the text and the word bank. If the participant is reading the whole text and only goes to the word bank a few times, because he has the ability to memorize the words, then this measure will also be affected by working memory. Our model goes along with the following conceptualization: both WM tests were related to the time spent in the text in relation to the word bank. In this case, the working memory span had a positive relationship with more time being spent on the text while the visuospatial working memory had an inverse proportional relationship.

Our models of the eye-tracking measures using the WM scores as predictors presented satisfactory results. The R^2 for the test data, predicting 30% of the variation of the number of toggles and 18% of the variation of the time-ratio on the text vs word bank, demonstrated reproducibility of the model. Furthermore, the cross-validated R^2 demonstrated that these variables in the test dataset could predict 1.6% and 0.3% of the variation in the number of toggles and in the time-ratio, respectively. These numbers may seem low, but the cross-validated R^2 is a pure indicator, which is different from the common R^2. Values from the model R^2 are fitted to the data and likely are inflated due to overfitting.

Another model which predicted the strategy applied based on the working memory scores and item was conducted. In this model, no relationship between the working memory span was found, while the visuospatial working memory was higher for those who use the cluster 2 strategy, the cluster related to a local strategy. This is a surprising result since previous literature (Khalifa & Weir, 2009; McCray & Brunfaut, 2018) suggested that participants with higher WM would do the global strategy, not the local strategy. Nevertheless, this result indicates that different components of WM may be applied in different strategies. In this case, participants with higher visuospatial WM may use it to try to understand the test based on a visual aspect instead of trying to comprehend the test. Therefore, this indicates that a local strategy is related to lower reading processes and balanced with other cognitive aspects, such as visuospatial abilities. It is noteworthy that the AUC of the ROC curve of the model in the test data was 0.731, while an ideal value would be over 0.750. Therefore, this model can be improved to achieve better reproducibility with other variables that go beyond the scope of this study.

The best performing model investigated the relationship between WM and the clusters with performance on the Cloze test. Since the lower IES scores present the best results, inversely proportional coefficients indicate a better result. In this case, when participants are doing a scanpath categorized as the strategy of Cluster 1, they perform better. Looking at the standardized beta, this was the biggest predictor of efficiency. This finding is very relevant because it indicates that the hypothesis of a global strategy being related to successfully achieving higher reading processes, mentioned by Khalifa and Weir (2009) and McCray and Brunfaut (2018), is confirmed. This happens because more proficient participants were able to be more efficient using a global strategy during the test. The study by McCray and Brunfaut (2018) demonstrated that lower performance participants tended to do a more local, lower-level processing. Additionally, it is interesting to note that, in more complex items, participants tended to use the local strategy. This indicates that when they had problems answering the item, they shifted to a local strategy, which is in line with Khalifa and Weir's (2009) hypothesis. Furthermore, both WM tests had a relationship with efficiency. In both cases, the higher the WM, the higher the efficiency. Therefore, our results indicate that both components of WM had a role in efficient reading. However, looking at the standardized beta, it is possible to see that the verbal WM played a bigger role in the efficiency in the cloze test.

It is noteworthy that the strategies found in our study are related to bigger concepts of reading abilities. This can be noted since the strategies changed according to the complexity of the test. In this sense, the strategies interplay with lower-order processes, such as decoding, and higher-order processes, such as the comprehension of idea units. Therefore, when a text is more complex, it can be due to word-level complexity. This can be verified by the use of less common words, which can compromise the decoding of the text. This problem can create a ripple effect until the text comprehension per se (Khalifa & Weir, 2009). Thus, it is probable that the strategies are also related to decoding, vocabulary, text comprehension, and other reading abilities.

Since the cloze gap-filling test is abundantly used for second language learning (e.g., Tremblay, 2011), the present study also has practical implications. Second language instructors and teachers should focus on teaching learners to use the global strategy, as this strategy is related to better performance. Therefore, teachers and instructors should teach students to read the whole text before starting to fill the gaps with words from the word bank. Furthermore, it is interesting to note that easier texts had predominantly test-takers doing global strategy, while intermediate and harder texts had the local strategy as predominant. This should be considered when creating items for the cloze test. Ideally, items should be of easy-to-intermediate difficulty, to enable participants to do both strategies. In these items, we can hypothesize that it will be easier to discriminate between good vs bad readers.

For future studies, we strongly encourage researchers to take these cognitive visual strategies into account. Furthermore, research should also focus on whether students struggle in the Cloze test due to bad employment of cognitive strategy. For example, is it possible that participants who are presenting a low performance can improve if they are taught how to apply the global strategy? Further studies should also research the relation between the strategies found in this study and other reading abilities, such as decoding, vocabulary, text comprehension, acoustic comprehension, verbal reasoning, among others.

To conclude, we found two visual cognitive strategies very similar to what was mentioned hypothesized in previous literature (Gao & Gu, 2008; McCray & Brunfaut, 2018; Yamashida, 2003). We found a relationship between these strategies with WM. Finally, we found that the strategies, verbal WM, and visuospatial WM were predictors of efficiency in the cloze test. In this sense, we were able to confirm that proficient participants employed a global strategy, indicating that is a relationship between the strategies employed in the task with higher-order processing.

Supplementary information

S1. Changes after preregistering the analysis.

In the AsPredicted, we mentioned that we would do an exploratory analysis to predict the performance in cloze test (last paragraph of Question 5). However, we decided to use a Ridge regularization (L2) instead of the Elastic-Net regularization (L1+L2). We decided to change it because we wanted to compare all the coefficients. Therefore, the exclusion of any of the independent variables by the L1 regularization would not be useful to us. Consequently, we decided to only use the L2 regularization. We implemented it solely on the model predicting efficiency and on the model predicting whether the participant answered the item correctly, that is only available in this Sup. File.

Outlier: Boxplot

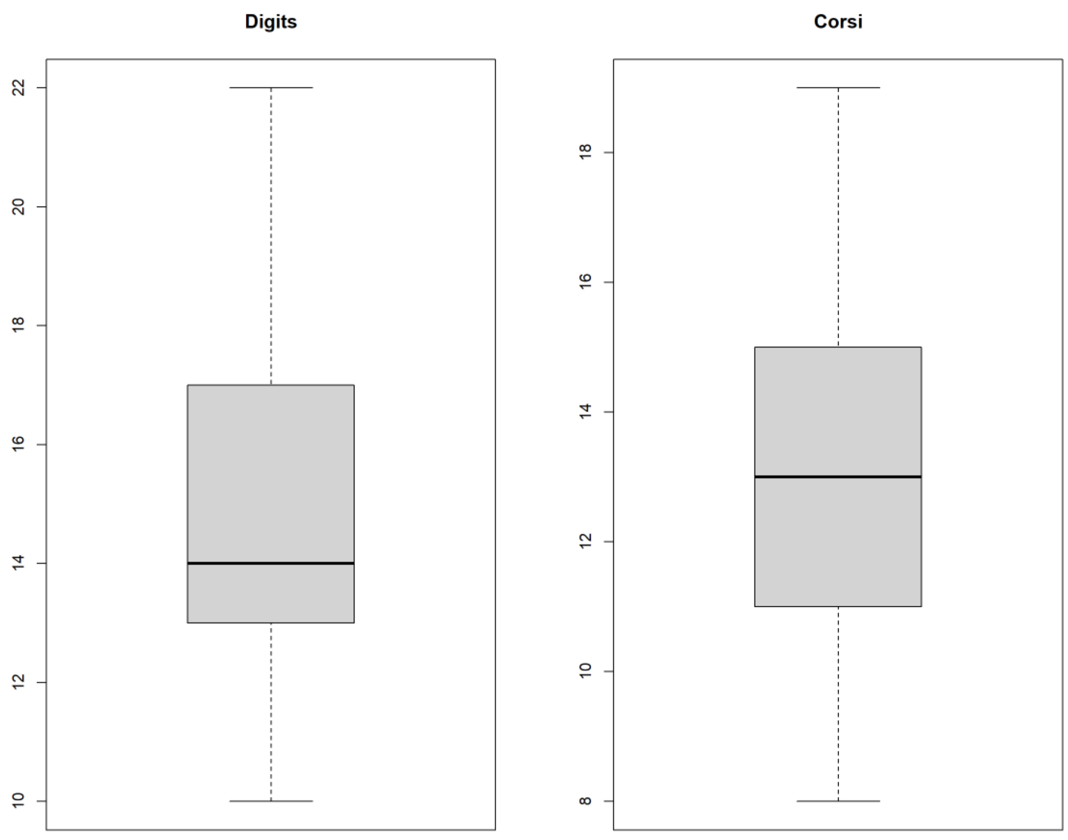

Figure S1. Boxplots of the total score of Digit Span task and the Corsi block-tapping test. Both boxplots do not present any outliers, which is confirmed by our analysis.

Scree plot from clusterization

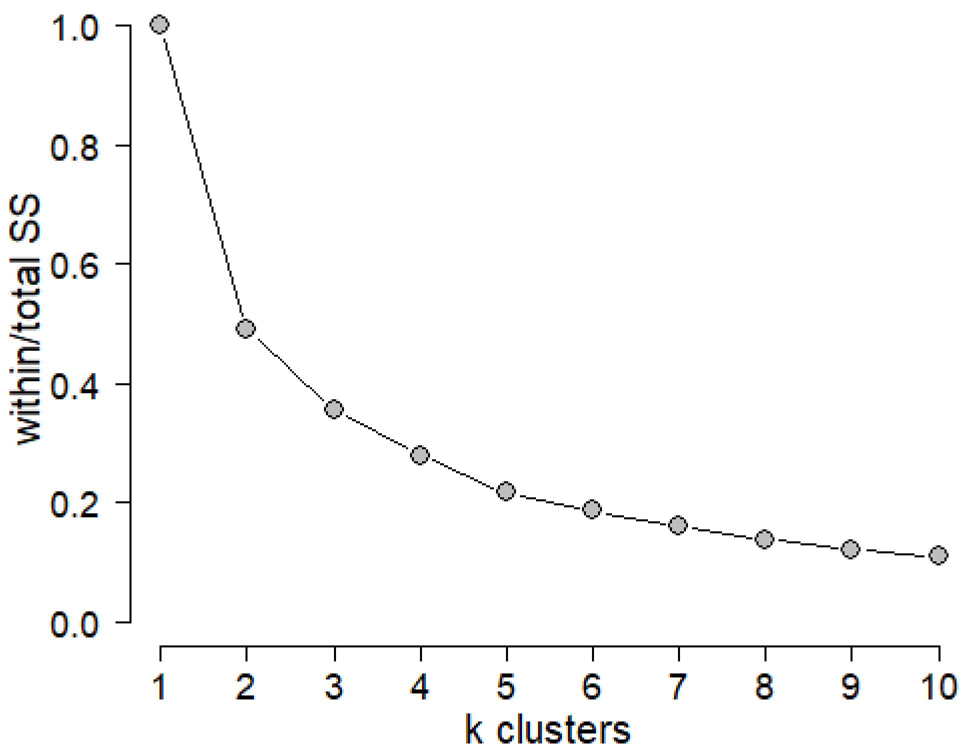

Figure S2. The scree plot of the clusterization model. In this plot, it is possible to see that the "elbow" of the plot is with 2 clusters. Furthermore, the difference of more than 3 clusters varies just a little. These results, while taking into account the previous literature whichmentions two strategies in the cloze test, made us decide to use k = 2.

Strategy used by each participant in each item

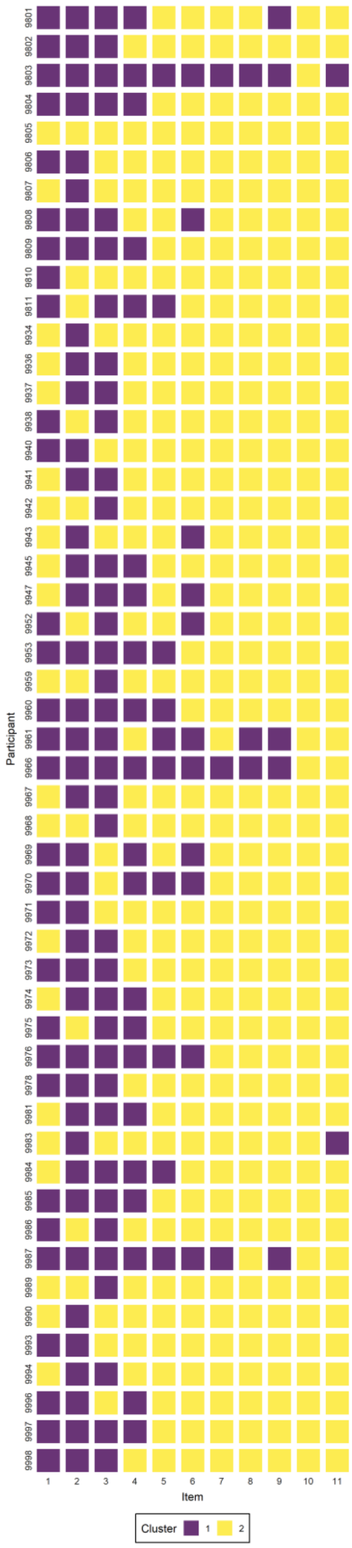

Figure S2. An infographic presenting which strategy each participant used in each item of the cloze test. Each column is an item completed by each individual, while the color of the square is the strategy that the participant used. Each row represents the set of items made by a participant.

Predicting if participants correctly answered an item of the cloze test

S2. The model predicting if the participant answered the cloze test item correctly.

We also conducted a model using the clusters, digit span, and Corsi block-tapping as the predictors in order to predict if the participants answered each item correctly. However, since there were only 9 incorrect trials, our elastic-net regression classified every trial as correct. This produced an accuracy of 91.89%, which was the same of the specificity, but no sensitivity could be observed. The coefficients of this regression are presented in Table S1.

Table S1. Coefficients and measures of the model predicting if the participant answered an item of the cloze test correctly.

		Correct or incorrect
Predictor's coefficients	Intercept	2.355
	Item	<-0.001
	Cluster 1	<0.001
	Digit Span	<0.001
	Corsi block-tapping	<0.001
Model measures	Alpha	1
	Lambda	100000
	Sensitivity	NA
	Specificity	0.9189

CONCLUSION

The main objective of this thesis was to analyze which cognitive strategies, measured by eye gaze patterns, are used in different tasks with answer banks, and to relate the cognitive strategies with several cognitive measures in order to pinpoint which cognitive abilities are related to these strategies. We tracked eye movements in a matrix reasoning task and a language proficiency gap-fill task and observed the relationship of these movements with different cognitive abilities. We were able to identify two strategies in the matrix reasoning task: constructive matching and response elimination. These strategies are present in the literature (e.g., Vigneau et al., 2006) and widely investigated. In the Cloze task, we found two strategies: a global and a local strategy. These strategies are very similar to the ones hypothesized in the literature (e.g., Gao & Gu, 2008; McCray & Brunfaut, 2018; Yamashida, 2003).

One of the first points to note is that working memory was related to eye movements in both tasks. This is not a surprising fact giving that working memory is the ability to manipulate different information (Baddeley, 2012). It is possible to imagine that the different quantity of information that a person can manipulate at once is related to how their eyes will behave while doing a complex task. For example, if a person has a small capacity in the working memory, they will gaze more times at different points of the task to remember the information that was presented there. On the other hand, if a person has a big capacity in the working memory, they will do a different eye gaze pattern: they will not do multiple comparisons since they have the information stored in their memory. This is very notable in how the working memory relates to alternances in matrix reasoning tasks, for example (Gonthier & Roulin, 2020).

The eye movement between two areas of interest, called “toggles” or “alternances” was a measure in both studies. In matrix reasoning tasks, this measure was related to better performance (Hayes et al., 2011, Laurence et al., 2018, Vigneau et al., 2006). In the gap-filling task, there is an indication that alternances are also related to good performance (McCray & Brunfaut, 2018). Our findings also suggest this idea. It is interesting to think about why the same eye movement behavior is related to different tasks. In the analogies task, Snow (1978, 1980), and Bethell-Fox and colleagues (1984) mentioned that participants could apply the response elimination strategy, where they would do multiple comparisons between the areas of interest in order to eliminate the wrong answers. Gonthier and Roulin (2020) also mention that this strategy is not effective and that test-takers rely on this strategy when their working memory capacity is not enough to process all the information in order to use a constructive matching strategy. Possibly, participants will do this eye movement of multiple comparisons when they do not have the cognitive capacity to do other eye gaze strategies effectively to answer correctly.

The studies presented here in this thesis have several practical implications in the long term. First, it is interesting to think that eye movements in a matrix reasoning task were able to predict executive functions in everyday life. Extrapolating these results, it might be possible to think that these eye movements have a lot to do with our daily performance. Second, these findings can impact the understanding of developmental disorders. For example, it is known that the eye movements in matrix reasoning tasks are different for Down syndrome and non-specific intellectual disability populations (Curie et al., 2016, Vakil & Lifshitz-Zehavi, 2012). Another developmental disorder that is very related to a non-typical eye gaze pattern is attention deficit hyperactivity disorder (Fried et al., 2014). Maybe, it is possible that visual training in different tasks can help attention deficit hyperactivity disorder patients. The development of cognitive abilities related to eye movements can transfer to other abilities and help with the attention deficit hyperactivity disorder symptoms. Obviously, this is a bold hypothesis, but new studies should focus on this direction.

Another direct consequence of this thesis is the knowledge related to how our eye movements are connected to different cognitive abilities. As eye tracking is becoming an increasingly common technology, with it even becoming possible via webcams (e.g., Finger et al., 2017), understanding how our eye movements are related to our cognitive abilities is important. The type of study presented in this thesis is a first step to contribute to the advance of this technology.

In this same direction, eye-tacking is becoming a valuable device for psychological assessment (e.g., Poletti et al., 2017). The knowledge of which cognitive process underlies the strategy use and the eye movements is a piece of essential information for developing psychological assessment via eye-tracking.

Several limitations can also be found in the studies. First, it would be much more interesting if there was only one sample with all measures collected from them. This would help to understand the relationships between several variables in a broader sense. Second, our samples consisted only of university students. Further studies should focus on different populations (e.g., children and teenagers, clinical groups), and on varied groups that these results can be generalized.

To summarize, we found that several cognitive abilities are related to different eye gaze patterns that reflect cognitive strategies. In particular, working memory seems to be an important component in these eye gaze patterns. Further, other abilities, such as planning, fluid reasoning, and executive functions also seem to have a connection with these eye movements.

REFERENCES

Alderson, J. C., & Cseresznyés, M. (2003). *Into Europe: Reading and use of English*. Budapest: Teleki László Foundation.

Alves, I. C. B. (2007). *Teste D.70: Manual revisado e ampliado* [D.70 Test: Revised and Extended Manual]. São Paulo, BR: Centro Editor de Testes e Pesquisas em Psicologia.

Anderson, J. R., Albert, M. V., & Fincham, J. M. (2005). Tracing problem solving in real time: fMRI analysis of the subject-paced Tower of Hanoi. *Journal of cognitive neuroscience*, *17*(8), 1261-1274. https://doi.org/10.1162/0898929055002427

Arendasy, M. E., & Sommer, M. (2013). Reducing response elimination strategies enhances the construct validity of figural matrices. *Intelligence*, *41*(4), 234-243. https://doi.org/10.1016/j.intell.2013.03.006

Ashcraft, M. H., & Fierman, B. A. (1982). Mental addition in third, fourth, and sixth graders. *Journal of Experimental Child Psychology*, *33*(2), 216-234. https://doi.org/10.1016/0022-0965(82)90017-0

Baddeley, A. (2012). Working memory: theories, models, and controversies. *Annual Review of Psychology*, *63*, 1-29.

Bailey, C. E. (2007). Cognitive accuracy and intelligent executive function in the brain and in business. *Annals of the New York Academy of Sciences*, *1118*(1), 122-141. https://doi.org/10.1196/annals.1412.011

Becker, N., Schmitz, F., Falk, A. M., Feldbrügge, J., Recktenwald, D. R., Wilhelm, O., & Spinath, F. M. (2016). Preventing response elimination strategies improves the convergent validity of figural matrices. *Journal of Intelligence*, *4*, 2. https://doi.org/10.3390/jintelligence4010002

Bereby-Meyer, Y., Assor, A., & Katz, I. (2004). Children’s choice strategies: The effects of age and task demands. *Cognitive Development*, *19*, 127-146. https://doi.org/10.1016/j.cogdev.2003.11.003

Bertolino, N. P. (2018). *Avaliação de compreensão de leitura: evidências a partir da análise dos movimentos oculares em teste do tipo Cloze* [Reading comprehension assessment: evidence from the analysis of eye movements in a Cloze test] (Unpublished Undergraduate research). Mackenzie Presbyterian University, São Paulo, Brazil.

Bethell-Fox, C. E., Lohman, D. F., & Snow, R. E. (1984). Adaptive reasoning: Componential and eye movement analysis of geometric analogy performance. *Intelligence*, *8*(3), 205-238. https://doi.org/10.1016/0160-2896(84)90009-6

Betsch, T., & Lang, A. (2013). Utilization of probabilistic cues in the presence of irrelevant information: A comparison of risky choice in children and adults. *Journal of Experimental Child Psychology*, *115*, 108-125. https://doi.org/10.1016/j.jecp.2012.11.003

Betta, E., & Turatto, M. (2006). Are you ready? I can tell by looking at your microsaccades. *Neuroreport*, *17*(10), 1001-1004. https://doi.org/10.1097/01.wnr.0000223392.82198.6d

Bexkens, A., Jansen, B. R., Van der Molen, M. W., & Huizenga, H. M. (2016). Cool Decision-Making in Adolescents with Behavior Disorder and/or Mild-to-Borderline Intellectual Disability. *Journal of abnormal child psychology, 44*(2), 357–367. https://doi.org/10.1007/s10802-015-9996-8

Bjorklund, D. F., & Rosenblum, K. E. (2001). Children's use of multiple and variable addition strategies in a game context. *Developmental Science, 4*(2), 184–194. https://doi.org/10.1111/1467-7687.00164

Borst, G., Poirel, N., Pineau, A., Cassotti, M., & Houdé, O. (2012). Inhibitory control in number-conservation and class-inclusion tasks: A neo-Piagetian inter-task priming study. *Cognitive Development*, *27*(3), 283-298. https://doi.org/10.1016/j.cogdev.2012.02.004

Brown, J. D. (2003). Twenty-five years of cloze testing research: so what? In Poedjosoedarmo, G. R. (Ed.), Teaching and assessing language proficiency (pp. 79–111). Singapore: SEAMEO Regional Language Centre.

Bruyer, R., & Brysbaert, M. (2011). Combining speed and accuracy in cognitive psychology: Is the inverse efficiency score (IES) a better dependent variable than the mean reaction time (RT) and the percentage of errors (PE)? *Psychologica Belgica*, *51*, 5–13. https://doi.org/10.5334/pb-51-1-5

Buczyłowska, D., Petermann, F., & Daseking, M. (2020). Executive functions and intelligence from the CHC theory perspective: Investigating the correspondence between the WAIS-IV and the NAB Executive Functions Module. *Journal of clinical and experimental neuropsychology*, *42*(3), 240-250. https://doi.org/10.1080/13803395.2019.1705250

Buswell, G. T. (1922). *Fundamental reading habits:* A study of their development.University of Chicago Press.

Buswell, G. T. (1937). *How adults read*. University of Chicago Press.

Buswell, G. T. (1935) *How people look at pictures*. University of Chicago Press.

Carroll, J. B. (1993). *Human cognitive abilities*. Cambridge: Cambridge University Press.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*, 1-22. https://doi.org/10.1037/h0046743.

Cherubino, P., Martinez-Levy, A. C., Caratu, M., Cartocci, G., Di Flumeri, G., Modica, E., Rossi, D., Mancini, M., & Trettel, A. (2019). Consumer behaviour through the eyes of neurophysiological measures: State-of-the-art and future trends. *Computational intelligence and neuroscience*, *2019*, 1976847. https://doi.org/10.1155/2019/1976847

Cho, S., Ryali, S., Geary, D. C., & Menon, V. (2011). How does a child solve 7+ 8? Decoding brain activity patterns associated with counting and retrieval strategies. *Developmental science*, *14*(5), 989-1001. https://doi.org/10.1111/j.1467-7687.2011.01055.x

Chuderski, A. (2013). When are fluid intelligence and working memory isomorphic and when are they not? *Intelligence*, *41*(4), 244-262. https://doi.org/10.1016/j.intell.2013.04.003

Clifton, C., Ferreira, F., Henderson, J. M., Inhoff, A. W., Liversedge, S. P., Reichle, E. D., & Schotter, E. R. (2016). Eye movements in reading and information processing: Keith Rayner's 40-year legacy. *Journal of Memory and Language*, *86*, 1-19. https://doi.org/10.1016/j.jml.2015.07.004

Corsi, P. M. (1972). Human memory and the medial temporal region of the brain. *Dissertation Abstracts International, 34*(2-B).

Cule, E., Moritz, S., & Frankowski, D. (2021). *ridge: Ridge Regression with Automatic Selection of the Penalty Parameter*. R Package version 2.9. https://CRAN.R-project.org/package=ridge

Curie, A., Brun, A., Cheylus, A., Reboul, A., Nazir, T., Bussy, G., ... & Marignier, S. (2016). A novel analog reasoning paradigm: new insights in intellectually disabled patients. *PloS one*, *11*(2), e0149717. https://doi.org/10.1371/journal.pone.0149717

Dehn, M. J. (2017). How working memory enables fluid reasoning. *Applied Neuropsychology: Child*, *6*(3), 245-247. https://doi.org/10.1080/21622965.2017.1317490

D'Esposito, M., & Postle, B. R. (2015). The cognitive neuroscience of working memory. *Annual review of psychology*, *66*, 115-142. https://doi.org/10.1146/annurev-psych-010814-015031

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

Drodick, L. W., Wahlstrom, D., Zhu, J., & Weiss, L. G. (2012). The Wechsler Adult Intelligence Scale—Fourth Edition and the Wechsler Memory Scale—Fourth Edition. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (p. 99–144). New York, NY: The Guilford Press.

Duchowski, A. T. (2007). *Eye tracking methodology*: Theory and Practice. Springer.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, *128*(3), 309–331. https://doi.org/10.1037/0096-3445.128.3.309

Figueiredo, V. L., & Nascimento, E. D. (2007). Desempenhos nas duas tarefas do subteste dígitos do WISC-III e do WAIS-III [Performance in the two digit span subtest tasks of the WISC-III and WAIS-III]. *Psicologia: teoria e pesquisa*, *23*(3), 313-318.

Finger, H., Goeke, C., Diekamp, D., Standvoß, K., & König, P. (2017). LabVanced: a unified JavaScript framework for online studies. In *International Conference on Computational Social Science (Cologne)*.

Flanagan, P., & Harrison, P. L. (2012), *Contemporary intellectual assessment: Theories, tests, and issues* (p. 99–144). New York, NY: The Guilford Press.

Fried, M., Tsitsiashvili, E., Bonneh, Y. S., Sterkin, A., Wygnanski-Jaffe, T., Epstein, T., & Polat, U. (2014). ADHD subjects fail to suppress eye blinks and microsaccades while anticipating visual stimuli but recover with medication. *Vision research*, *101*, 62-72. https://doi.org/10.1016/j.visres.2014.05.004

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, *33*(1), 1-22.

Gao, X., & Gu, X. (2008). An introspective study on test-taking process for banked cloze. *CELEA Journal*, *31*(4), 3-16.

Gathercole, S. E., & Pickering, S. J. (2000). Working memory deficits in children with low achievements in the national curriculum at 7 years of age. *British Journal of Educational Psychology*, *70*(2), 177-194. https://doi.org/10.1348/000709900158047

Gathercole, S. E., Durling, E., Evans, M., Jeffcock, S., & Stone, S. (2008). Working memory abilities and children's performance in laboratory analogues of classroom activities. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *22*(8), 1019-1037. https://doi.org/10.1002/acp.1407

Gilchrist, I. D. (2011). Saccades. In S. P. Liversedge, I. D. Gilchrist, S. Everling. *The Oxford Handbook of Eye Movements* (pp. 85-94). Oxford University Press.

Gonthier, C., & Roulin, J.-L. (2020). Intraindividual strategy shifts in Raven’s matrices, and their dependence on working memory capacity and need for cognition. *Journal of Experimental Psychology: General*, *149*(3), 564–579. https://doi.org/10.1037/xge0000660

Gonthier, C., & Thomassin, N. (2015). Strategy use fully mediates the relationship between working memory capacity and performance on Raven’s matrices. *Journal of Experimental Psychology: General*, *144*(5), 916–924. https://doi.org/10.1037/xge0000101

Gottfredson, L. S. (1997a). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence, 24*, 13-23. https://doi.org/10.1016/S0160-2896(97)90011-8

Gottfredson, L. S. (1997b). Why g matters: The complexity of everyday life. *Intelligence*, *24*(1), 79-132.

Greene, M. R., Liu, T., Wolfe, J. M. (2012). Reconsidering Yarbus: A failure to predict observers’ task from eye movement patterns. *Vision research*, *62*, 1-8. https://doi.org/10.1016/j.visres.2012.03.019

Groner, R., & Groner, M. (1982). Towards a hypothetico-deductive theory of cognitive activity. In R. Groner, P. Fraisse. *Cognition and eye movements* (pp.100-121). North Holland. https://www.researchgate.net/publication/312424385

Hayes, T. R., Petrov, A. A., & Sederberg, P. B. (2011). A novel method for analyzing sequential eye movements reveals strategic influence on Raven's Advanced Progressive Matrices. *Journal of Vision*, *11*(10), 10-10. https://doi.org/10.1167/11.10.10

Hayes, T. R., Petrov, A. A., & Sederberg, P. B. (2015). Do we really become smarter when our fluid-intelligence test scores improve? *Intelligence*, *48*, 1-14. https://doi.org/10.1016/j.intell.2014.10.005

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtiss, G. (2004). *Teste Wisconsin de Classificação de Cartas* [Wisconsin Card Sorting Test]. São Paulo, BR: Casa do Psicólogo.

Hering, E. (1879). Über Muskelgeräusche des Auges [About eye muscle noises]. *Sitzberichte der kaiserlichen Akademie der Wissenschaften in Wien. Mathematischnaturwissenschaftliche Klasse*, *79*, 137-154.

Hessels, M. G., Vanderlinden, K., & Rojas, H. (2011). Training effects in dynamic assessment: A pilot study of eye movement as indicator of problem solving behaviour before and after training. *Educational and Child Psychology*, *28*(2), 101.

Hodgson, T. L., Bajwa, A., Owen, A. M., & Kennard, C. (2000). The strategic control of gaze direction in the Tower of London task. *Journal of Cognitive Neuroscience*, *12*(5), 894–907. https://doi.org/10.1162/089892900562499

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford: Oxford University Press.

Huey, E. B. (1898). Preliminary experiments in the physiology and psychology of reading. *The American Journal of Psychology*, *9*(4), 575-586.

Huey, E. B. (1908). *The psychology and pedagogy of reading*. Macmillan.

Huizenga, H. M., Crone, E. A., & Jansen, B. J. (2007). Decision-making in healthy children, adolescents and adults explained by the use of increasingly complex proportional reasoning rules. *Developmental science*, *10*(6), 814-825. https://doi.org/10.1111/j.1467-7687.2007.00621.x

Ispas, D., & Borman, W. C. (2015). Personnel selection, psychology of. In J. D. Wright (Ed.), *International encyclopedia of the social & behavioral sciences* (2nd ed., p. 937). Oxford, England: Elsevier.

Ito, J., Yamane, Y., Suzuki, M., Maldonado, P., Fujita, I., Tamura, H., & Grün, S. (2017). Switch from ambient to focal processing mode explains the dynamics of free viewing eye movements. *Scientific reports*, *7*, 1082. https://doi.org/10.1038/s41598-017-01076-w

Jana, T. A. (2018). *Análise do padrão dos movimentos oculares em uma multitarefa de busca visual em ambiente real e suas relações com as funções executivas* [Analysis of the pattern of eye movements in a visual search multitasking in a real environment and its relationships with executive functions] (Unpublished master's thesis). Mackenzie Presbyterian University, São Paulo, Brazil.

Jansen, B. R. J, & van der Maas, H. L. J. (2001). Evidence for the phase transition from Rule I to Rule II on the balance scale task. *Developmental Review, 21*(4), 450-494. https://doi.org/10.1006/drev.2001.0530

Jansen, B. R. J, & van der Maas, H. L. J. (2002). The development of children's rule use on the balance scale task. *Journal of Experimental Child Psychology*, *81*(4), 383-416. https://doi.org/10.1006/jecp.2002.2664

Jarosz, A. F., & Wiley, J. (2012). Why does working memory capacity predict RAPM performance? A possible role of distraction. *Intelligence*, *40*(5), 427-438. https://doi.org/10.1016/j.intell.2012.06.001

Jarosz, A. F., Raden, M. J., & Wiley, J. (2019). Working memory capacity and strategy use on the RAPM. *Intelligence*, *77*, 101387. https://doi.org/10.1016/j.intell.2019.101387

Jastrzębski, J., Ciechanowska, I., & Chuderski, A. (2018). The strong link between fluid intelligence and working memory cannot be explained away by strategy use. *Intelligence*, *66*, 44-53. https://doi.org/10.1016/j.intell.2017.11.002

Jastrzębski, J., Ociepka, M., & Chuderski, A. (2020). Fluid reasoning is equivalent to relation processing. *Intelligence*, *82*, 101489. https://doi.org/10.1016/j.intell.2020.101489

Javal, E. (1878). Essai sur la physiologie de la lecture [Essay on the Physiology of Reading]. *Annales d'Ocilistique, 80*, 240-274.

Javal, (1879). E. Essai sur la physiologie de la lecture [Essay on the Physiology of Reading]. *Annales d'Ocilistique*, *82*, 242-253.

Jeffreys, H. (1961). *Theory of Probability*. Oxford: Oxford University Press

Judd, C. H, Mcallister, C. N., & Steel, W.M. (1905). General introduction to a series of studies of eye movements by means of kinetoscopic photographs. In J. M. Baldwin, H. C. Warren, & C. H. Judd. *Psychological Reviews* (pp.1-16). Monograph supplements.

Kasneci, E., Kasneci, G., Trautwein, U., Appel, T., Tibus, M., Jaeggi, S. M., & Gerjets, P. (2021). Do your eye movements reveal your performance on an IQ test? A study linking eye movements and socio-demographic information to fluid intelligence. *PsyArXiv*. https://doi.org/10.31234/osf.io/dru93

Kaufman, S. B. (2014). *Working Memory and Fluid Reasoning: Same or Different?* Scientific American. https://blogs.scientificamerican.com/beautiful-minds/working-memory-and-fluid-reasoning-same-or-different/

Khalifa, H., & Weir, C. J. (2009). *Examining reading: Research and practice in assessing second language reading*. Cambridge: Cambridge University Press.

Kleijn, S., Pander Maat, H., & Sanders, T. (2019). Cloze testing for comprehension assessment: The HyTeC-cloze. *Language Testing*, *36*(4), 553-572. https://doi.org/10.1177%2F0265532219840382

Krikorian, R., Bartok, J., & Gay, N. (1994). Tower of London procedure: a standard method and developmental data. *Journal of clinical and Experimental Neuropsychology*, *16*(6), 840-850. https://doi.org/10.1080/01688639408402697

Kucharský, Š., Visser, I., Truțescu, G. O., Laurence, P. G., Zaharieva, M., & Raijmakers, M. E. (2020). Cognitive strategies revealed by clustering eye movement transitions. *Journal of Eye Movement Research*, *13*. https://doi.org/10.16910%2Fjemr.13.1.1

Kwak, Y., Payne, J. W., Cohen, A. L., & Huettel, S. A. (2015). The rational adolescent: Strategic information processing during decision making revealed by eye tracking. *Cognitive development*, *36*, 20-30. https://doi.org/10.1016/j.cogdev.2015.08.001

Lamare, M. (1892). Des mouvements des yeux dans la lecture [About the eye movements in reading]. *Bulletins et mémoires de la Société française d'Ophthalmologie*, *10*, 354-364.

Laurence, P. G., Mecca, T. P., Serpa, A., Martin, R., & Macedo, E. C. (2018). Eye movements and cognitive strategy in a fluid intelligence test: Item type analysis. *Frontiers in psychology*, *9*, 380. https://doi.org/10.3389/fpsyg.2018.00380

Lindner, A., Iyer, A., Kagan, I., & Andersen, R. A. (2010). Human posterior parietal cortex plans where to reach and what to avoid. *Journal of Neuroscience*, *30*(35), 11715-11725. https://doi.org/10.1523/JNEUROSCI.2849-09.2010

Loesche, P., Wiley, J., & Hasselhorn, M. (2015). How knowing the rules affects solving the Raven Advanced Progressive Matrices Test. *Intelligence*, *48*, 58-75. https://doi.org/10.1016/j.intell.2014.10.004

Luke, S. G., Darowski, E. S. & Gale, S. D. (2018) Predicting eye-movement characteristics across multiple tasks from working memory and executive control. *Memory & Cognition* 46**,** 826–839. https://doi.org/10.3758/s13421-018-0798-4

Macgregor, J. N., Ormerod, T. C., & Chronicle, E. P. (2001). Information processing and insight: a process model of performance on the nine-dot and related problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 176-201. https://psycnet.apa.org/doi/10.1037/0278-7393.27.1.176

Martinez-Conde, S., & Macknik, S. L. (2011). Microsaccades. In S. P. Liversedge, I. D. Gilchrist, & S. Everling. *The Oxford Handbook of Eye Movements* (pp. 95-114). Oxford University Press.

Martinez-Conde, S., Macknik, S. L., Hubel, D. H. (2004). The role of fixational eye movements in visual perception. *Nature reviews neuroscience*, *5*(3), 229-240. https://doi.org/10.1038/nrn1348

McCray, G., & Brunfaut, T. (2018). Investigating the construct measured by banked gap-fill items: Evidence from eye-tracking. *Language Testing*, *35*, 51-73. https://doi.org/10.1177%2F0265532216677105

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence, 37*, 1–10. https://doi.org/10.1016/j.intell.2008.08.004

Meghanathan, R. N., van Leeuwen, C., & Nikolaev, A. R. (2015). Fixation duration surpasses pupil size as a measure of memory load in free viewing. *Frontiers in Human Neuroscience,* 8. https://doi.org/10.3389/fnhum.2014.01063

Merchant, J., Morrissette, R., & Porterfield, J. L. (1974). Remote measurement of eye direction allowing subject motion over one cubic foot of space. *IEEE transactions on biomedical engineering*, (4), 309-317. https://doi.org/10.1109/TBME.1974.324318

Mitchum, A. L., & Kelley, C. M. (2010). Solve the problem first: Constructive solution strategies can influence the accuracy of retrospective confidence judgments. *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, *36*(3), 699–710. https://doi.org/10.1037/a0019182

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current directions in psychological science*, *21*, 8-14. https://doi.org/10.1177%2F0963721411429458

Mogle, J. A., Lovett, B. J., Stawski, R. S., & Sliwinski, M. J. (2008). What's so special about working memory? An examination of the relationships among working memory, secondary memory, and fluid intelligence. *Psychological science*, *19*(11), 1071-1077. https://doi.org/10.1111/j.1467-9280.2008.02202.x

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Prentice-Hall.

Oller, J. W., & Jonz, J. (1994). Why cloze Procedure? In Oller, J. W., Jonz, J. (Eds.), *Cloze and coherence*. Cranbury, NJ: Bucknell University Press.

Ones, D. S., Dilchert, S., Viswesvaran, C., (2012). Cognitive abilities. In: Schmitt, N. (Ed.), *Oxford Handbook of Personnel Assessment and Selection*. Oxford UniversityPress, New York, pp. 179–224

Otero-Millan, J., Troncoso, X. G., Macknik, S. L., Serrano-Pedraza, I., & Martinez-Conde, S. (2008). Saccades and microsaccades during visual fixation, exploration, and search: foundations for a common saccadic generator. *Journal of vision*, *8*(14), 21-21. https://doi.org/10.1167/8.14.21

Papagiannopoulou, E. A., Chitty, K. M., Hermens, D. F., Hickie, I. B., & Lagopoulos, J. (2014). A systematic review and meta-analysis of eye-tracking studies in children with autism spectrum disorders. *Social neuroscience*, *9*(6), 610-632. https://doi.org/10.1080/17470919.2014.934966

Pieters, R., & Wedel, M. (2012). Ad gist: Ad communication in a single eye fixation. *Marketing Science*, *31*(1), 59-73. https://doi.org/10.1287/mksc.1110.0673

Płużyczka, M. (2018). The First Hundred Years: a History of Eye Tracking as a Research Method. *Applied Linguistics Papers, 45*(4), 101-116. http://doi.org/10.32612/uw.25449354.2018.4.pp.101-116

Poirel, N., Borst, G., Simon, G., Rossi, S., Cassotti, M., Pineau, A., & Houdé, O. (2012). Number conservation is related to children’s prefrontal inhibitory control: an fMRI study of a Piagetian task. *PloS one*, *7*(7), e40802. https://doi.org/10.1371/journal.pone.0040802

Poletti, B., Carelli, L., Solca, F., Lafronza, A., Pedroli, E., Faini, A., ... & Silani, V. (2017). An eye-tracking controlled neuropsychological battery for cognitive assessment in neurological diseases. *Neurological Sciences*, *38*(4), 595-603. https://doi.org/10.1007/s10072-016-2807-3

Rabi, R., Miles, S. J., & Minda, J. P. (2015). Learning categories via rules and similarity: Comparing adults and children. *Journal of experimental child psychology*, *131*, 149-169. https://doi.org/10.1016/j.jecp.2014.10.007

Ratcliff, R., Thapar, A., & McKoon, G. (2011). Effects of aging and IQ on item and associative memory. *Journal of Experimental Psychology: General*, *140*(3), 464. https://dx.doi.org/10.1037/a0023810

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

Rayner, K. (1979). Eye guidance in reading: Fixation locations within words. *Perception*, *8*, 21-30. https://doi.org/10.1068%2Fp080021

Rayner, K. (1978). Eye movements in reading and information processing. *Psychological Bulletin, 85*(3), 618–660. https://doi.org/10.1037/0033-2909.85.3.618

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. https://doi.org/10.1037/0033-2909.124.3.372

Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive psychology*, *7*, 65-81. https://doi.org/10.1016/0010-0285(75)90005-5

Rivollier, G., Quinton, J. C., Gonthier, C., & Smeding, A. (2020). Looking with the (computer) mouse: How to unveil problem-solving strategies in matrix reasoning without eye-tracking. Behavior Research Methods, 1-16. https://doi.org/10.3758/s13428-020-01484-3

Rosa, A. T. F. (2016). *Análise do padrão dos movimentos oculares em tarefas de leitura: relações entre processamento lexical, sintático e semântico* [Analysis of eye movement pattern in reading tasks: relationships between lexical, syntactic and semantic processing] (Unpublished doctoral dissertation). Mackenzie Presbyterian University, São Paulo, Brazil.

Roth, R. M., Isquith, P. K., & Goia, G. A. (2005). *Behavior Rating Inventory of Executive Functions – Adult Version*. Lutz, FL: Psychological Assessment Resources

Roth, R. M., Lance, C. E., Isquith, P. K., Fischer, A. S., & Giancola, P. R. (2013). Confirmatory factor analysis of the behavior rating inventory of executive function-adult version in healthy adults and application to attention-deficit/hyperactivity disorder. *Archives of clinical neuropsychology*, *28*(5), 425-434. https://doi.org/10.1093/arclin/act031

Santos, F. H., Mello, C. B., Bueno, O. F. A., & Dellatolas, G. (2005). Cross-cultural differences for three visual memory tasks in Brazilian children. *Perceptual and motor skills*, *101*(2), 421-433. https://doi.org/10.2466/pms.101.2.421-433

Schlottfeldt, C. G., & Malloy-Diniz, L. (2018). *Teste Matrizes de Vienna 2: Versão Informatizada. [Vienesse Matrices Test 2: Informatized Version]*. São Paulo, BR: Editora Hogrefe Cetepp.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological bulletin*, *124*(2), 262. https://doi.org/10.1037/0033-2909.124.2.262

Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (p. 99–144). New York, NY: The Guilford Press.

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *298*(1089), 199-209. https://doi.org/10.1098/rstb.1982.0082

Siegler, R. S. (2007). Cognitive variability. *Developmental science*, 10, 104-109. https://doi.org/10.1111/j.1467-7687.2007.00571.x

Siegler, R. S. (1988). Individual differences in strategy choices: Good students, not-so-good students, and perfectionists. *Child development*, *59*(4), 833-851.

Siegler, R. S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General*, *116*(3), 250–264. https://doi.org/10.1037/0096-3445.116.3.250

Snow, R. E. (1978). Eye Fixation and Strategy Analyses of Individual Differences in Cognitive Aptitudes. In A. M. Legold, J. W. Pellegrino, S. D. Fokkema & R. Glaser. *Cognitive psychology and instruction* (p. 299-308). New York, NY: Plenum Press.

Snow, R. E. (1980). Aptitude processes. In R. E. Snow, P.-A. Federick & W. E. Montague. *Aptitude learning, and instruction: Cognitive process analysis of aptitude* (p. 27-63). Hillsdale, NJ: Erlbaum.

Starr, A., Vendetti, M. S., & Bunge, S. A. (2018). Eye movements provide insight into individual differences in children's analogical reasoning strategies. *Acta Psychologica*, *186*, 18-26. https://doi.org/10.1016/j.actpsy.2018.04.002

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, *87*(2), 245–251. https://doi.org/10.1037/0033-2909.87.2.245

Steingroever, H., Jepma, M., Lee, M. D., Jansen, B. R., & Huizenga, H. M. (2019). Detecting strategies in developmental psychology. *Computational Brain & Behavior*, *2*(2), 128-140. https://doi.org/10.1007/s42113-019-0024-x

Sweller, J., Levine, M. (1982). Effects of goal specificity on means–ends analysis and learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*(5), 463-474. https://doi.org/10.1037/0278-7393.8.5.463

Tatler, B. W., Wade, N. J., Kwan, H., Findlay, J. M., & Velichkovsky, B. M. (2010). Yarbus, eye movements, and vision. *i-Perception*, *1*, 7-27. https://doi.org/10.1068/i0382

Thibaut, J. P., & French, R. M. (2016). Analogical reasoning, control and executive functions: a developmental investigation with eye-tracking. *Cognitive Development*, *38*, 10-26. https://doi.org/10.1016/j.cogdev.2015.12.002

Thompson, V. A. (2021). Eye-tracking IQ: Cognitive capacity and strategy use on a ratio-bias task. *Cognition*, *208,* 104523. https://doi.org/10.1016/j.cognition.2020.104523

Torbeyns, J., De Smedt, B., Ghesquière, P., & Verschaffel, L. (2009). Acquisition and use of shortcut strategies by traditionally schooled children. *Educational Studies in Mathematics*, *71*, 1-17. https://doi.org/10.1007/s10649-008-9155-z

Tremblay, A. (2011). Proficiency assessment standards in second language acquisition research:" Clozing" the gap. *Studies in Second Language Acquisition*, 339-372. https://doi.org/10.1017/S0272263111000015

Unsworth, N., Spillers, G. J., & Brewer, G. A. (2009). Examining the relations among working memory capacity, attention control, and fluid intelligence from a dual-component framework. *Psychology Science, 51*(4), 388–402.

Vabalas, A., Gowen, E., Poliakoff, E., & Casson, A. J. (2019). Machine learning algorithm validation with a limited sample size. *PloS one*, *14*(11), e0224365. https://doi.org/10.1371/journal.pone.0224365

Vakil, E., & Lifshitz-Zehavi, H. (2012). Solving the Raven Progressive Matrices by adults with intellectual disability with/without Down syndrome: Different cognitive patterns as indicated by eye-movements. *Research in Developmental Disabilities*, *33*(2), 645-654. https://doi.org/10.1016/j.ridd.2011.11.009

Vakil, E., Lifshitz, H., Tzuriel, D., Weiss, I., & Arzuoan, Y. (2011). Analogies solving by individuals with and without intellectual disability: Different cognitive patterns as indicated by eye movements. *Research in Developmental Disabilities*, *32*(2), 846-856. https://doi.org/10.1016/j.ridd.2010.08.006

Valsecchi, M., Betta, E., & Turatto, M. (2007). Visual oddballs induce prolonged microsaccadic inhibition. *Experimental Brain Research*, *177*(2), 196-208. https://doi.org/10.1007/s00221-006-0665-6

Valsecchi, M., & Turatto, M. (2009). Microsaccadic responses in a bimodal oddball task. *Psychological research*, *73*, 23-33. https://doi.org/10.1007/s00426-008-0142-x

van Aken, L., van der Heijden, P. T., Oomens, W., Kessels, R. P., & Egger, J. I. (2019). Predictive value of traditional measures of executive function on broad abilities of the cattell–horn–carroll theory of cognitive abilities. *Assessment*, *26*(7), 1375-1385. https://doi.org/10.1177/1073191117731814

Van Der Maas, H. L., Molenaar, D., Maris, G., Kievit, R. A., & Borsboom, D. (2011). Cognitive psychology meets psychometric theory: On the relation between process models for decision making and latent variable models for individual differences. *Psychological review*, *118*(2), 339-356. https://doi.org/10.1037/a0022749

Van Der Maas, H. L., & Molenaar, P. C. (1992) Stagewise cognitive development: an application of catastrophe theory. *Psychological review*, *99*(3), 395-417. https://doi.org/10.1037/0033-295X.99.3.395

Vendetti, M. S., Starr, A., Johnson, E. L., Modavi, K., & Bunge, S. A. (2017). Eye movements reveal optimal strategies for analogical reasoning. *Frontiers in psychology*, *8*, 932. https://doi.org/10.3389/fpsyg.2017.00932

Vigneau, F., Caissie, A. F., & Bors, D. A. (2006). Eye-movement analysis demonstrates strategic influences on intelligence. *Intelligence, 34*(3), 261–272. https://doi.org/10.1016/j.intell.2005.11.003

Vinegrad, M. (1994). A revised adult dyslexia check list. *Educare*, *48*, 21–23.

Wade, N. J. (2010). Pioneers of eye movement research. *i-Perception*, *1*(2), 33-68.

Wade, N. J., & Tatler, B. W. (2009). Did Javal measure eye movements during reading? *Journal of Eye Movement Research*, *2*(5). https://doi.org/10.16910/jemr.2.5.5

Wade, N. J., Tatler, B. W., Heller, D. (2003). Dodge-ing the issue: Dodge, Javal, Hering, and the measurement of saccades in eye-movement research. *Perception*, *32*(7), 793-804. https://doi.org/10.1068%2Fp3470

Wechsler, D. (2004). *WAIS-III*. São Paulo, BR: Casa do Psicólogo.
Alderson, J. C. (2000). *Assessing reading*. Cambridge: Cambridge University Press.

Yamashita, J. (2003). Processes of taking a gap-filling test: comparison of skilled and less skilled EFL readers. *Language testing*, *20*(3), 267-293. https://doi.org/10.1191%2F0265532203lt257oa

Yarbus, A. L. (1967). *Eye movements and vision*. Plenum Press.