Intelligence 100 (2023) 101782

Contents lists available at ScienceDirect

Intelligence

journal homepage: www.elsevier.com/locate/intell

Using a multi-strategy eye-tracking psychometric model to measure intelligence and identify cognitive strategy in Raven's advanced progressive matrices

Yaohui Liu [a], Peida Zhan [a,b,c,*], Yanbin Fu [a], Qipeng Chen [a], Kaiwen Man [d], Yikun Luo [a]

[a] School of Psychology, Zhejiang Normal University, Jinhua City 321004, China
[b] Intelligent Laboratory of Child and Adolescent Mental Health and Crisis Intervention of Zhejiang Province, Jinhua 321004, China
[c] Key Laboratory of Intelligent Education Technology and Application of Zhejiang Province, Jinhua 321004, China
[d] Educational Research Program, Educational Studies in Psychology, Research Methodology, and Counseling, University of Alabama, Tuscaloosa, AL, USA

ARTICLE INFO

Keywords:
Intelligence
Cognitive strategy
Eye movement
Item response theory
Rasch model

ABSTRACT

Previous studies have found that participants use two cognitive strategies—constructive matching and response elimination—in responding to items in the Raven's Advanced Progressive Matrices (APM). This study proposed a multi-strategy psychometric model that builds on item responses and also incorporates eye-tracking measures, including but not limited to the proportional time on matrix area (PTM), the rate of toggling (ROT), and the rate of latency to first toggle (RLT). By jointly analyzing item responses and eye-tracking measures, this model can measure each participant's intelligence and identify the cognitive strategy used by each participant for each item in the APM. Several main findings were revealed from an eye-tracking-based APM study using the proposed model: (1) The effects of PTM and RLT on the constructive matching strategy selection probability were positive and higher for the former than the latter, while the effect of ROT was negligible. (2) The average intelligence of participants who used the constructive matching strategy was higher than that of participants who used the response elimination strategy, and participants with higher intelligence were more likely to use the constructive matching strategy. (3) High-intelligence participants increased their use of the constructive matching strategy as item difficulty increased, whereas low-intelligence participants decreased their use as item difficulty increased. (4) Participants took significantly less time using the constructive matching strategy than the response elimination strategy. Overall, the proposed model follows the theory-driven modeling logic and provides a new way of studying cognitive strategy in the APM by presenting quantitative results.

1. Introduction

General intelligence is a core component of the intelligence structure and is considered to be an important predictor of academic and professional success, which has been discussed for decades (Kane, Hambrick, & Conway, 2005; Marshalek, Lohman, & Snow, 1983; Vigneau, Caissie, & Bors, 2006). Raven's Advanced Progressive Matrices (APM; Raven, Raven, & Court, 1998) is a standardized cognitive ability test designed to measure general intelligence or the aptitude to solve new problems by some elementary cognitive processes, such as identifying relations, drawing inferences (Loesche, 2020; McGrew, 2009). More specifically, as a strictly visual test, APM is one of the most commonly used measurement instruments for fluid intelligence than of crystallized intelligence (Loesche, 2020). Further, the APM has been widely used in research on cognitive strategies (Laurence, Mecca, Serpa, Martin, & Macedo, 2018) that represent sets of cognitive processes in the process of solving problems or achieving goals (Cameron & Cameron & Jago, 2013; Lemaire & Reder, 1999).

Fig. 1 (Left) displays an item from the APM. The item consists of a three-by-three matrix with figural elements in the matrix area and eight options in the response options area. One of the cells in the matrix area is missing and needs to be selected from the response options area, which requires participants to understand the rules hidden in the matrix area to make the selection (Gonthier, 2022). Two common cognitive strategies—constructive matching and response elimination—were revealed in previous eye-tracking-based studies for participants to solve items in the APM (Vigneau et al., 2006). In constructive matching, participants first construct a mental representation of the answer and then make a

* Corresponding author at: Room 731, Building 17, School of Psychology, Zhejiang Normal University, Jinhua City 321004, Zhejiang Province, China.
E-mail address: zhan@zjnu.edu.cn (P. Zhan).

https://doi.org/10.1016/j.intell.2023.101782
Received 20 December 2022; Received in revised form 3 July 2023; Accepted 4 July 2023
Available online 18 July 2023

choice, whereas in response elimination, participants make a choice by comparing the elements of the matrix with different options multiple times.

The application of eye-tracking technology enhances our exploration and understanding of cognitive processes (Wedel, Pieters, & van der Lans, 2022). Previous studies have found a correspondence between the cognitive strategies used by participants in the APM and their visual search patterns inferred from eye-tracking measures (e.g., Vigneau et al., 2006). Specifically, compared with participants who use the response elimination strategy, participants who use the constructive matching strategy typically spend proportionately more time in the matrix area than in the response area, display a smaller number of toggles between the matrix and response options areas, and take a longer time to look from the matrix area to the response options area for the first time. Further, when using the constructive matching strategy, participants have a lower toggling rate, which indicates the number of toggles between the matrix and response options areas per second, but they toggle at a higher rate when using the response elimination strategy (Laurence et al., 2018). Although these studies explored the relationship between eye movements and cognitive strategies, they did not provide a specific cognitive strategy identification method.

The identification of cognitive strategies is the basis for exploring the relationship between cognitive strategy and other cognitive factors (e.g., cognitive style and working memory), as well as for cognitive strategy training (e.g., Hayes, Petrov, & Sederberg, 2015; Hessels, Vanderlinden, & Rojas, 2011; Jarosz & Wiley, 2012; Kucharský et al., 2020; Li, Ren, Schweizer, & Wang, 2022). Previous studies attempted to identify participants' cognitive strategies using one of two self-reporting approaches. The first approach uses self-report questionnaires to ask participants to report the cognitive strategies they use to solve problems throughout the entire APM (Gonthier & Thomassin, 2015; Jastrzębski, Ciechanowska, & Chuderski, 2018; Li et al., 2022; Mitchum & Kelley, 2010). The main limitation of this approach is that individuals may not be able to accurately and objectively recall the strategies they use, and the approach cannot ascertain whether participants switch strategies throughout the test (Jarosz, Raden, & Wiley, 2019; Lemaire & Reder, 1999). The second approach involves utilizing the think-aloud protocol to ask participants to continuously say their thoughts aloud at each step during problem solving (Jarosz et al., 2019). However, participants may not be able to decipher all of their thoughts completely and accurately during problem solving, and the unstructured nature of language makes it difficult to accurately identify the cognitive strategies they employ. Moreover, verbal protocols can increase the cognitive load, cause the participant to be unable to express well, or might influence the participant's response process or task performance (Chiu & Shu, 2010; Jarosz et al., 2019).

Further, considering that different strategies may correspond to different visual search patterns (Thibaut & French, 2016; Vigneau et al., 2006), some researchers have also tried to use data-driven methods to distinguish different strategies from eye-tracking data. For example, to analyze eye-tracking data (e.g., saccade), Hayes, Petrov, and Sederberg (2011) proposed a semi-supervised algorithm (i.e., successor representation scanpath analysis), which combines a higher-order probability transfer matrix and a Markov model to visualize participants' visual search patterns. Although this algorithm can accurately predict participants' test scores in the APM and can provide insight into the differences in their problem solving, it is rarely used in other studies due to its computational complexity, the subjective nature of the interpretation of results, and the ambiguity of strategy identification (Hayes et al., 2015; Laurence, 2021). Kucharský et al. (2020) proposed an unsupervised algorithm for mining eye-tracking data, which combines probability transfer matrix and K-means clustering; however, they found that the visual search patterns obtained from clustering by this algorithm did not match well with theoretically existing cognitive strategies, probably due to the limited amount of data. Overall, the currently used eye-tracking data mining algorithms are mainly limited by the low interpretability of the results and high data volume requirements.

In general, because eye-tracking technology can capture participants' eye movements in a detailed and objective manner, eye-tracking-based cognitive strategy identification methods are theoretically more accurate and objective than self-reporting approaches. Given the limits of existing eye-tracking data mining approaches, we need a novel eye-tracking-based cognitive strategy identification method with high interpretability of results and appropriately small data volume requirements.

This study aimed to incorporate eye-tracking measures into the

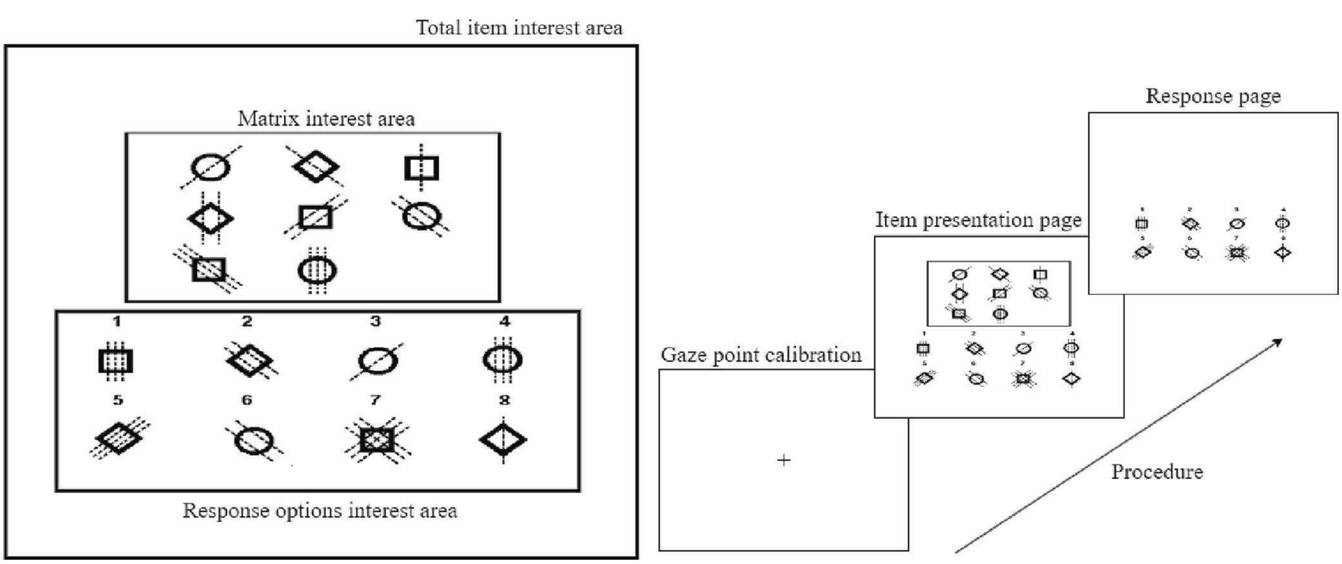

Fig. 1. Example of the Raven item format and test sequence.
Note. (Left) The item can be divided into three interest areas: the total item interest area, the matrix interest area, and the response options interest area. The item consists of a three-by-three matrix with figural elements in the matrix area and eight options in the response options area. One of the cells in the matrix area is missing and needs to be selected from the response options. (Right) Each test has three phases: gaze point calibration, item presentation, and response. Eye movements are collected during the item presentation phase. Pressing the spacebar on the keyboard to enter the response page, only the options are presented on the screen, and the participant presses one of the number keys 1–8 on the keyboard to respond.

cognitive strategy identification method in the APM from the perspective of psychometric modeling to propose a multi-strategy eye-tracking model (denoted as MEM). By jointly analyzing item responses and eye-tracking measures, the proposed model can measure each participant's intelligence that is reflected by the latent trait and identify the cognitive strategy used by each participant for each item in the APM. Theoretically, compared with eye-tracking data mining approaches, the proposed model is more interpretable and has fewer data volume requirements, which makes it more suitable for small-scale psychological experimental studies. Further, compared with the self-reporting approaches, the proposed model can more objectively identify participants' cognitive strategies for each item. In the rest of the paper, we illustrate the performance of the proposed model and the corresponding findings from an eye-tracking-based APM study.

2. Methods

2.1. Instrument

The APM contains 48 homogenous items, which are divided into two sets: the first set of 12 items is intended to be used as a practice set to familiarize participants with the format, and the second set of 36 items is the main test that is supposed to be scored. Typically, the first set is omitted or reduced to one or two example items (e.g., Carpenter, Just, & Shell, 1990; Dehn, 2017; Hayes et al., 2015; Kpolovie & Emekene, 2016). In this study, to prevent participants' practice of the first set of items from affecting their responses to the second set of items used in the formal experiment, we used three easy items from Raven's Standard Progressive Matrices as warm-up items to familiarize participants with the response process.

2.2. Participants

A total of 202 students participated in this study. All were randomly selected from a university in a coastal province in China and had not participated in the APM before. A total of 10 students were excluded, 7 of whom had serious missing or unexplainable gaze points on many items (see Appendix A), and 3 had test scores that were more than three standard deviations below the average of all students' scores (M = 25.12, SD = 4.92). The results from the final 192 participants (147 females and 45 males; $M_{\text{age}} = 22.06$, $SD = 2.54$) were used for further analysis.

2.3. Procedure

Before the test began, the participants were informed of the entire test process. Each participant who completed the test carefully received a cash payment. The test program was carried out with *E*-Prime software (version 2.0) to record the outcome data, including item response accuracy and item response times[1]; Eye-tracking data were collected with a non-contact device, SR Research Ltd. EyeLink Portable Duo (EyeLink Portable Duo - SR Research (sr-research.com)), set in monocular tracking mode with a sampling rate of 1000 HZ. Each participant was placed in a quiet room, sitting approximately 60 cm from the computer screen, and chin support was used to assist in fixing the participant's head. A staff member operated the eye-tracking device not far away to ensure the smooth running of the experiment (see Appendix B for the experimental environment).

Fig. 1 (Right) displays the test sequence. The participants were given nine warm-up items (three items to be exact, each repeated three times) to familiarize them with the purpose and procedure of the test before the formal test began. In the calibration session, a 5-point calibration method was used to calibrate the participant's gaze point before allowing the participant to enter the response procedure. In the response procedure inspired by Hayes et al. (2011), a cross mark was first presented at the center of the screen. The participant was required to look at the cross mark; if the gaze point deviation was larger than 2 cm, the participant returned to the calibration session; otherwise, they moved forward to the next item presentation page. On the item presentation page, there was no time limit for the participants to reason the answer. When they were ready to respond, the participant pressed the spacebar on the keyboard to enter the response page, where only the options were presented on the screen; pressing one of the number keys 1–8 on the keyboard allowed them to respond. Furthermore, considering that the participants may experience visual fatigue during the whole test, we allowed them to take a short break, usually less than one minute, when the test was halfway through.

2.4. Interest areas and eye-tracking measures

Eye-tracking techniques were used to analyze participants' strategies in the APM, and specific eye-tracking measures depended on the division of interest areas. In this study, three interest areas (i.e., total item interest area, matrix interest area, and response options interest area) were defined (see Fig. 1(Left)). Eye movements beyond the interest areas were ignored.

Following existing studies on cognitive strategies and visual search patterns, three composite eye-tracking measures were recorded and calculated in this study (Hayes et al., 2011; Laurence et al., 2018; Li et al., 2022; Vigneau et al., 2006): the proportional time on matrix area (PTM), the rate of toggling (ROT), and the rate of latency to first toggle (RLT).

2.4.1. Item latency and proportional time on the matrix area

Item latency (T^{item}) is the time that participants spend on the total item interest area, which is recorded as the time that participants spend on the item presentation page (i.e., item response time). The proportional time on the matrix of each item was calculated from the time spent on the matrix interest area (T^{matrix}) divided by the time spent on the item interest area; that is, $PTM = T^{matrix}/T^{item}$; the higher the value, the higher proportion of time a participant spent in the matrix interest area.

2.4.2. Number of toggles and rate of toggling

A shift of gaze from one area to another is considered a toggle. The rate of toggling was calculated by dividing the item latency (T^{item}) by the number of toggles (G); that is, $ROT = T^{item}/G$ (Laurence, 2021),[2] which is more suitable for representing the strategies compared with the number of toggles (Laurence, 2021) and reflects how long it takes to transition once between the matrix interest area and the response interest area; the higher the value, the lower the number of transitions of a participant's gaze within the same time.

2.4.3. The latency to first toggle and the rate of latency to first toggle

The latency to first toggle is also an important measure, which is the time it takes participants to transfer their gaze point from the matrix interest area to options for the first time (Vigneau et al., 2006). The rate of latency to first toggle is equal to the latency to first toggle (T^{toggle})

[1] According to the description of the study procedure below, the item response time only includes the time spent by the participants in the item presentation page but not the time spent by the participants in the response page. The advantage of this procedure is that it eliminates the time spent by the participant on the keystroke response and reduces potential error (e.g., Hayes et al., 2011).

[2] In some previous studies (e.g., Laurence et al., 2018; Vigneau et al., 2006), ROT was calculated as ROT = G / T^{item}, indicating how many toggles per second, which is the inverse of the ROT used in this study.

divided by the item latency (T^{item}); that is, $RLT = T^{toggle}/T^{item}$; the higher the value, the higher proportion of time a participant spent on planning solutions to the problem. Compared with T^{toggle}, RLT additionally takes into account the processing speed of individuals, as reflected by T^{item}. Suppose T^{toggle} of two participants are both 5 s, one participant's T^{item} is 10 s, and the other's T^{item} is 20 s. The latency to first toggle is 50% of the response time for the first participant and 25% of that for the second participant, which indicates that it would not be appropriate to assume that both participants use the same strategy.

Based on the findings of previous studies (e.g., Hayes et al., 2011; Laurence et al., 2018; Vigneau et al., 2006), we assumed that participants who use the constructive matching strategy would have a higher RLT, a higher ROT, and a higher PTM than those who use the response elimination strategy, indicating that they spend more time analyzing the matrix before the first gaze at the options, take longer to each toggle between the matrix and the options, and spend more time on the matrix area throughout the entire response process.

2.5. Multi-strategy psychometric modeling incorporating eye-tracking measures

Psychometric modeling is a theory-driven approach that constructs the probabilistic relationship between observed data and latent variables through statistical distribution functions. Currently, several multi-strategy psychometric models have been proposed to identify students' problem-solving strategies in the field of educational achievement tests, such as mixture or multigroup models (e.g., Mislevy & Verhelst, 1990; von Davier, 2010) and multi-strategy cognitive diagnostic models (e.g., de La Torre & Douglas, 2008; Ma & Guo, 2019). One major limitation of mixture or multigroup models is that participants are usually assumed to use one of multiple strategies for all items in a test. Although multi-strategy cognitive diagnostic models eliminate this non-switching strategy assumption, these models are still limited to traditional item response accuracy data and categorial latent variables, ignoring the process of problem solving that can be reflected by eye movements. Perhaps due to the lack of communication between researchers in the field of intelligence and psychometricians, to our knowledge, no studies have attempted to use multi-strategy psychometric models to identify participants' cognitive strategies in the APM, let alone involving eye-tracking measures.

In recent years, a few studies have combined eye-tracking measures and psychometric models to assess participants' visual engagement in educational assessments or to assess participants' decision-making behavior in naturalistic settings (e.g., Man & Harring, 2019; Wedel et al., 2022; Zhan, Man, Wind, & Malone, 2022). Although these studies did not involve the identification of cognitive strategy, their idea of incorporating eye-tracking measures into psychometric models inspired the current study to construct the MEM, namely, a multi-strategy psychometric model involving eye-tracking measures (e.g., PTM, ROT, and RLT).

2.5.1. Generalized modeling and assumptions

Essentially, the MEM is primarily designed to infer the probability of cognitive strategy use based on eye-tracking measures, which are then combined with item response accuracy data to estimate the participant's intelligence.

Consider participants taking an intelligence test with J items and M predetermined cognitive strategies. Based on the total probability theorem, the proposed MEM can be expressed as:

$$P\left(Y_{ij}=1|\theta_i\right)=\sum_{m=1}^{M}P\left(Y_{ij}=1|\theta_i,m_{ij}\right)\times P\left(m_{ij}\right), \tag{1}$$

where $P\left(Y_{ij}=1|\theta_i\right)$ denotes the correct response probability of participant i (i = 1, 2, …, I) to item j (j = 1, 2, …, J), given the participant's latent trait (θ_i) used to reflect the intelligence. $P\left(Y_{ij}=1|\theta_i,m_{ij}\right)$ is the strategy implementation model, which denotes the correct response probability of participant i to item j, given the participant's intelligence θ_i and the cognitive strategy m_{ij} (m_{ij} = 1, 2, …, M) he/she used to item j. The proposed model allows participants to switch their cognitive strategies across items.

Considering the limited number of participants in eye-tracking studies (usually just a few dozen), a Rasch family model is used to define the strategy implementation model:

$$P\left(Y_{ij}=1|\theta_i,m_{ij}=m\right)=\frac{exp\left(\theta_i-b_j+e_{mj}\right)}{1+exp\left(\theta_i-b_j+e_{mj}\right)}, \tag{2}$$

where b_j denotes the difficulty of item j, and e_{mj} is the strategy sensitivity parameter, which represents the variation in the correct response probability by using strategy m on item j; the larger the value of $|e_{mj}|$, the more sensitive item j is to strategy m.

$P\left(m_{ij}\right)$ is the strategy selection model, and it denotes the probability of participant i applying strategy m on item j, and $\sum_{m=1}^{M}P\left(m_{ij}\right)=1$. It is inferred using eye-tracking measures in the present study, and its value is constrained to a number between 0 and 1, with a logistic function as follows:

$$P\left(m_{ij}=m\right)=\frac{exp\left(\omega_1\times f_{1ij}+\omega_2\times f_{2ij}+\ldots+\omega_c\times f_{cij}\right)}{1+exp\left(\omega_1\times f_{1ij}+\omega_2\times f_{2ij}+\ldots+\omega_c\times f_{cij}\right)}, \tag{3}$$

where f_{cij} represents the c-th (c = *1, 2, …, C*) eye-tracking measure of participant i on item j, and ω_c is the corresponding weight parameter of it. The weight parameters reflect the magnitude of the effect of different eye-tracking measures on the probability of strategy selection. Note that in the strategy selection model, each eye-tracking measure for all participants can be standardized for each item to put all weight parameters on the same scale. This allows for a comparison of the extent of the impact of various eye-tracking measures on the probability of strategy selection.

Based on the model setup, some assumptions of the MEM can be summarized for a better understanding of it. First, following the findings of existing studies on strategy choice in cognitive tasks (e.g., the probability-matching perspective and the over-matching perspective [e.g., Lovett & Anderson, 1995; Ma & Guo, 2020]), the MEM assumes that a participant will use different cognitive strategies in proportion to their responses simultaneously, rather than definitively using a specific strategy. Second, during problem-solving, unlike the ability, the choice and use of strategies does not directly determine the success or failure of the problem solving, but affects the process and time spent on problem solving (Cook & Mayer, 1983; Mislevy et al., 1991); hence, the MEM assumes that participants' choice of cognitive strategy is not theoretically related to their intelligence but is reflected by behavioral indicators of the problem-solving process (i.e., eye-tracking measures in this study).[3] Third, the MEM assumes that the impact of cognitive strategy on the correct response probability is reflected in the item parameter via the strategy sensitivity parameter. In other words, the impact on the correct response probability is the same for participants using the same cognitive strategy on a given item; of course, this assumption could be released as $e_{mj}\Longrightarrow e_{mij}$, but this would undoubtedly increase the complexity of the model and increase the demand on the amount of data volume.

2.5.2. Specific setting

Some specific settings can be made to the generalized function of the MEM to apply to the analysis of APM data. In this study, $M = 2$ cognitive

[3] Note that since no latent trait parameter was involved in the strategy selection model in the MEM, the intelligence reflected by latent trait and the choice of strategy are theoretically independent, but this does not mean that they are statistically independent.

strategies were predetermined, namely, the constructive matching strategy and the response elimination strategy. Specifically, $m_{ij} = 1$ indicates that participant i applied the constructive matching strategy to item j, and $m_{ij} = 2$ indicates that participant i applied the response elimination strategy to item j.

Prior studies generally accept that the constructive matching strategy is a more effective strategy with higher accuracy or test scores relative to the response elimination strategy (Arendasy & Sommer, 2013; Gonthier & Roulin, 2020; Gonthier & Thomassin, 2015; Mitchum & Kelley, 2010). Therefore, we assumed that the correct response probability for participants who use the constructive matching strategy would be higher than or equal to that for those who use the response elimination strategy, namely:

$$\mathrm{P}\left(Y_{ij} = 1|\theta_i, m_{ij} = 1\right) \geq \mathrm{P}\left(Y_{ij} = 1|\theta_i, m_{ij} = 2\right). \quad (4)$$

More specifically,

$$\mathrm{P}\left(Y_{ij} = 1|\theta_i, m_{ij} = 1\right) = \frac{exp\left(\theta_i - b_j + e_{1j}\right)}{1 + exp\left(\theta_i - b_j + e_{1j}\right)} = \frac{exp\left(\theta_i - b_j + e_j\right)}{1 + exp\left(\theta_i - b_j + e_j\right)}, \quad (5)$$

$$P\left(Y_{ij} = 1|\theta_i, m_{ij} = 2\right) = \frac{exp\left(\theta_i - b_j + e_{2j}\right)}{1 + exp\left(\theta_i - b_j + e_{2j}\right)} = \frac{exp\left(\theta_i - b_j\right)}{1 + exp\left(\theta_i - b_j\right)}, \quad (6)$$

where $e_{1j} = e_j$ and is constrained to be non-negative (i.e., $e_j \geq 0$), and $e_{2j} = 0$; in such cases, the larger e_j is, the greater the gain in the correct response probability of item j for participants who use the constructive matching strategy compared to those who use the response elimination strategy. Also, we can determine whether participants can obtain an improvement in correct response probability by using the constructive matching strategy based on whether the item strategy sensitivity parameter is significantly greater than zero.

Three eye-tracking measures (i.e., PTM, ROT, and RLT) were used in the strategy selection model, as follows:

$$P\left(m_{ij} = 1\right) = \frac{exp\left(\omega_1 \times f_{1ij} + \omega_2 \times f_{2ij} + \omega_3 \times f_{3ij}\right)}{1 + exp\left(\omega_1 \times f_{1ij} + \omega_2 \times f_{2ij} + \omega_3 \times f_{3ij}\right)}, \quad (7)$$

and

$$P\left(m_{ij} = 2\right) = 1 - P\left(m_{ij} = 1\right), \quad (8)$$

where f_{1ij}, f_{2ij}, and f_{3ij} represent participant i's PTM, ROT, and RLT on item j, respectively. Given the positive weight parameters, the higher the value of the three eye-tracking measures, the participants are more likely to adopt the constructive matching strategy (Laurence, 2021; Li et al., 2022). In summary, Eqs. (1) and (4–8) together constitute the MEM used in this study.

2.6. Analysis

SR Research Ltd. EyeLink Data Viewer (Data Viewer for EyeLink Eye-Tracker Gaze Data – SR Research [sr-research.com]) was used to analyze and export the recorded eye-tracking data. Eye-tracking measures were calculated for 192 participants, and the mean-fill method was used to handle the missing values (<1%) that occurred on a few items. We calculated the eye-tracking measures for each participant on each item; further, the three eye-tracking measures for all participants were standardized for each item to put the weight parameters on the same scale.

The parameters of the MEM and its sub-models can be estimated using the Bayesian Markov Chain Monte Carlo (MCMC) algorithm via Just Another Gibbs Sampler (JAGS) software (Version 4.3.0; Plummer, 2015). The process of parameter estimation was performed based on Python software (Version 3.10.6). The JAGS code with prior distributions of all model parameters and MCMC procedure are provided in Appendix C. To increase the repeatability of the current study, all relevant data and the Python running code used in this study are available at https://osf.io/wx2p8/?view_only=9218af49196e4eb3bb947ba68f8f66c8. More details about how JAGS is used for Bayesian estimation can be found in Zhan, Jiao, Man, and Wang (2019).

To present the advantages of considering the cognitive strategy, we compared the fit of the MEM and the Rasch model (Rasch, 1960), which does not consider the cognitive strategy to the APM data.[4] The widely available information criterion (WAIC) and leave-one-out cross-validation (LOO) (Gelman et al., 2014; Vehtari, Gelman, & Gabry, 2016) were used as the relative model-data fit indices; smaller values indicate a better model–data fit. Further, posterior predictive model checking (PPMC) (Gelman et al., 2014) was used to evaluate the absolute model–data fit; a posterior predictive probability (*ppp*) value near 0.5 indicates that there are no systematic differences between the predictive and observed data and thus an adequate fit of the model; by contrast, when the *ppp* value smaller than 0.025 or larger than 0.975 indicates the model does not fit the data. In this study, the differences between the observed data, $\boldsymbol{Y}$, and posterior predicted data, $\boldsymbol{Y}^{postpred}$, were compared in computing the PPMC; that is, $ppp = \sum_{e=1}^{E}\left(sum\left(\boldsymbol{Y}^{postpred(e)}\right) \geq sum(\boldsymbol{Y})\right)/E$, where E is the total number of iterations in MCMC sampling; $\boldsymbol{y}^{postpred(e)}$ indicates the posterior predicted data in the e-th iteration, which were generated from the item response function (e.g., Eq. (1) of the MEM) based on the samplings of the model parameters from the posterior distributions.

The validity of the identification results of the model was verified by manual judgment. First, two staff members who had not participated in the study were trained and informed of the definition of the two cognitive strategies and the corresponding typical eye movements. Second, the eye-tracking diagrams (heat map and gaze plot) of five participants were randomly selected from the eye-tracking diagrams of 192 participants for each item, and the corresponding model identification results were extracted.[5] Then, two staff members were independently asked to make their judgment of whether to endorse the model identification results based on the eye-tracking diagrams (see Appendix D). Lastly, we assessed the consistency of the two staff members' judgment results.

3. Results

3.1. Main results

Table 1 summarizes the absolute and relative model–data fit indices of the MEM and the Rasch model. Both models fit the APM data well in

Table 1
Summaries of absolutely and relatively model–data fit indices.

Analysis model	*ppp*	WAIC	LOO
Rasch model	0.44	6485.43	6486.57
MEM	0.44	6395.67	6397.08

Note. MEM: multi-strategy eye-tracking model; ppp: posterior predictive probability; WAIC: widely available information criterion; LOO: leave-one-out cross-validation.

[4] The Rasch model has been used in some previous studies to analysis the APM data (e.g., Waschl, Nettelbeck, Jackson, & Burns, 2016) and can be expressed as $P\left(Y_{ij} = 1|\theta_i\right) = \frac{exp\left(\theta_i - b_j\right)}{1 + exp\left(\theta_i - b_j\right)}$, where θ_i is latent trait (i.e., intelligence) of participant i and b_j is the difficulty of item j.

[5] In this study, to identify the strategy used by participants, the strategy selection probabilities for all participants to all items were dichotomized according to a cut-point of 0.5. Specifically, the participants were identified as using the constructive matching strategy on item j by $P\left(m_{ij} = 1\right)\rangle 0.5$ and the response elimination strategy on item j by $P\left(m_{ij} = 1\right) \leq 0.5$.

terms of the *ppp* value, indicating that their analysis results can be used to reflect the characteristics implied by the data. However, the MEM fit the data better than the Rasch model according to the WAIC and LOO, indicating that additional consideration of cognitive strategy better reflected the characteristics of the data. Fig. 2 displays the scatterplot of intelligence (i.e., latent trait) estimates for the two models ($r = 0.986$, $p < 0.001$); such a high correlation indicated that the two models measured the same latent trait, namely, the additional consideration of cognitive strategy did not change the latent trait measured by the model. The following sections will focus on the analysis results of the MEM.

Fig. 3 displays the scatterplot of participants' raw scores and latent trait estimates of the MEM ($r = 0.984$, $p < 0.001$). Although there is a high positive correlation between raw scores and intelligence estimates, the two are not equivalent. The MEM can further differentiate the intelligence of participants who received the same raw score. Such advantage of the MEM comes from the theoretical advantages of item response theory over classical test theory (Embretson & Reise, 2013; Rasch, 1960). Specifically, instead of classifying participants into 37 categories (i.e., 0–36 scores) only, as in the case of raw scores, the MEM can make full use of the participant's response information and separate the effect of item characteristics (e.g., difficulty) and participant's latent trait on the response, thus achieving a more refined measurement of the participant's intelligence (i.e., for participants with the same raw score, the higher the difficulty of the correctly responded items, the higher the participants' latent trait estimate is likely to be). The following discussions will use the intelligence reflected by the latent trait of the MEM.

Figs. E1–E3 in Appendix E successively present the distribution of the three eye-tracking measures—PTM, ROT, and RLT—on 36 items. The estimated weight coefficients of the three eye-tracking measures were $\omega_1 = 2.01$ (95%highest posterior density $[\text{HPD}] = [0.79, 3, 42]$) for PTM,[6] $\omega_2 = 0.08$ ($95\%\text{HPD} = [-0.77, 0.94]$) for ROT, and $\omega_3 = 1.32$($95\%\text{HPD} = [0.64, 2.12]$) for RLT, respectively, indicating that the predictions of constructive matching strategy selection probability of PTM and RLT were positive, while ROT seemed redundant. To further investigate the effect of different eye-tracking measures on strategy choice, six sub-models of the MEM were additionally used to analyze the APM data, including three sub-models containing any two of the three eye-tracking measures in Eq. (7) (denoted as MEM2) and three sub-models containing any one of the three eye-tracking measures in Eq. (7) (denoted as MEM1). Table E1 in Appendix E presents the estimated weight parameters and two relative model–data fit indices for the seven models. The results showed that the weight parameter of ROT was not equal to zero only for the worst-fitting MEM1 that contained ROT alone; the weight parameter of ROT in the other models was not significantly different from zero. The MEM2 with ROT removed on the basis of MEM fit the data best, according to two relatively model–data fit indices. As shown in Table E2 in Appendix E, the MEM and the MEM2 without ROT had the highest consistency in the identification of cognitive strategies among the seven models. Overall, PTM, RLT were in descending order of importance for the constructive matching strategy selection probability, and ROT had no significant effect in MEM.

Fig. 4 displays the item difficulty estimates and strategy sensitivity estimates of the 36 items in the APM. First, there was a tendency for the items to increase in difficulty as the test progressed, the Spearman rank correlation coefficient between item difficulty and item serial number is 0.853 ($p < 0.001$). Second, the strategy sensitivity parameter varied by item, indicating that participants had a greater relative advantage in responding to some items (e.g., items 9, 16, 21, 22, and 36) using the constructive matching strategy than using the response elimination strategy, whereas the impact of strategy use was small in responding to some other items (e.g., items 1, 5, 8, 10, and 13). Strategy sensitivity parameter estimates for 19 items did not differ significantly from zero, indicating that using the constructive matching strategy on these items may not bring a gain to their correct response probability.

Fig. 5 displays the Spearman rank correlation coefficients among five indicators, including (a) the mean constructive matching strategy selection probability per item ($CMSSP_j = \sum_{i=1}^{I} P(m_{ij})/I$), (b) the mean response time across participants per item ($MRT_j = \sum_{i=1}^{I} T_{ij}^{item}/I$), (c) the difference between the mean abilities of two strategy groups per item ($DMA_j = \sum_{i=1}^{I}\left(\theta_i^{high} - \theta_i^{low}\right)\Big/I$), (d) item difficulty b_j, and (e) strategy sensitivity e_j.[7] First, there was a significantly high positive correlation between b_j and MRT_j, indicating that the higher the item difficulty, the longer the time participants took to respond. Second, a significantly moderate positive correlation was found between e_j and DMA_j, indicating that the higher item strategy sensitivity, the greater the difference between the mean abilities of the two strategy groups. Third, there was a marginal significance moderate positive correlation between b_j and DMA_j, which seemed to indicate that the higher the item difficulty, the greater the difference between the mean abilities of the two strategy groups. Lastly, there was no significant correlation between CMSSP_j and any of the other four indicators, indicating that for all participants, the use of the constructive matching strategy was not significantly influenced by item difficulty, strategy sensitivity, average time spent, or the difference in ability between the two strategy groups of participants. However, some previous studies (e.g., Bethell-Fox, Lohman, & Snow, 1984; Jarosz et al., 2019) found that participants' strategy use may interact with their intelligence and the difficulty of items (e.g., low intelligence participants tend to use response elimination strategy on difficult items). Thus, it is necessary to explore further the relationship between participants' intelligence, item difficulty, and constructive matching strategy selection probability.

To this end, we conducted a linear regression analysis with constructive matching strategy selection probability $P(m_{ij})$ as the dependent variable and intelligence θ_i, item difficulty b_j, and their interaction ($\theta_i \times b_j$) as independent variables.[8] Table 2 presents the results of the regression analysis. The regression effect was significant ($F = 29.36$; $p < 0.001$). In addition, the main effect of intelligence was significantly positive, indicating participants with higher intelligence were more likely to use the constructive matching strategy (Fig. E4 displays the scatterplot of all participants' intelligence estimates and their mean constructive matching strategy selection probability across all items). The main effect of item difficulty was non-significant, consistent with the result in Fig. 5. More importantly, the interaction term was significant, indicating that the level of intelligence affects the relationship between item difficulty and constructive matching strategy selection probability.

To further demonstrate the effects of intelligence on the relationship between item difficulty and constructive matching strategy selection probability, we divided the participants into high-, medium-, and low-intelligence groups according to whether they fell into the top 30%, middle 40%, and bottom 30% of the intelligence estimates (Beuchert & Mendoza, 1979; Engelhart, 1965). The mean constructive matching strategy selection probability for all items was 55.7% for the high-intelligence group, 50.4% for medium-intelligence group, and 44.5% for the low-intelligence group. As shown in Fig. 6, we calculated the

[6] The function of the 95% highest posterior density in Bayesian statistics is similar to that of the 95% confidence interval in frequentist statistics when its range contains zero, indicating that the estimate (i.e., the posterior mean) is not significantly different from zero.

[7] Some indicators do not meet the assumption of normal distribution required for Pearson correlation.

[8] We uniformly convert all variables into vectors of length $I \times J$, where the values of three independent variables paired with $P(m_{ij})$ are b_j, θ_i, and $\theta_i \times b_j$, respectively.

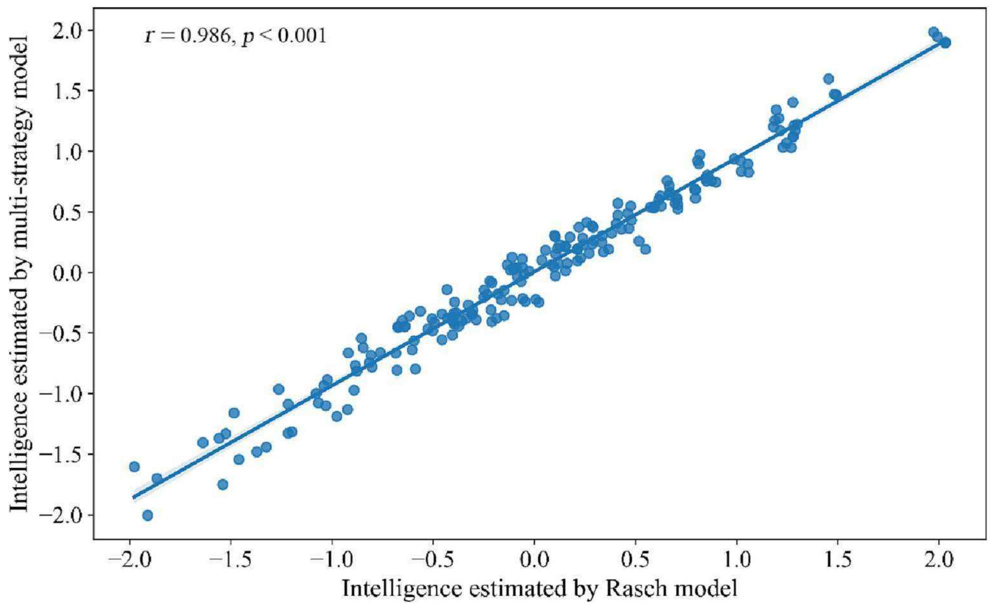

Fig. 2. Scatterplot of intelligence estimates for the Rasch model and multi-strategy eye-tracking model.

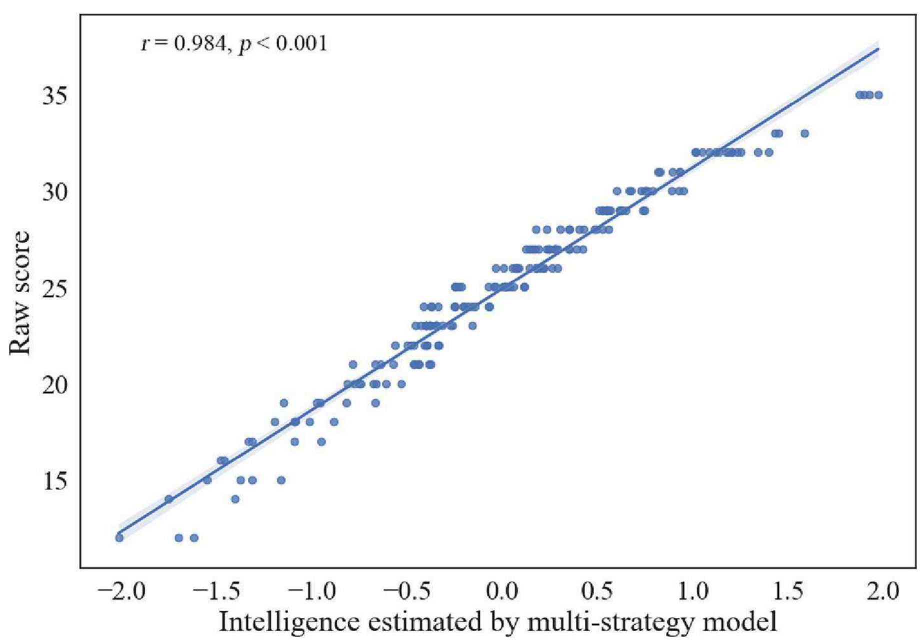

Fig. 3. Scatterplot of raw scores and intelligence estimates of the multi-strategy eye-tracking model.

Pearson correlation coefficient between the constructive matching strategy selection strategy per item (i.e., $CMSSP_j$) and item difficulty in three groups, respectively.[9] There was a significant moderate positive correlation for the high-intelligence group $r(CMSSP_j, b_j) = 0.493$ ($p = 0.002$), a non-significant low negative correlation for the medium-intelligence group $r(CMSSP_j, b_j) = -0.269$ ($p = 0.113$), and a marginally significant low negative correlation for the low-intelligence group $r(CMSSP_j, b_j) = -0.298$ ($p = 0.078$). Such results indicated that participants in the high-intelligence group preferred the constructive matching strategy on difficulty items; in contrast, participants in the low-intelligence group decreased their use of the constructive matching strategy as item difficulty increases.

Fig. 7 displays nine eye-tracking diagrams (heat maps and gaze plots) of three participants on three APM items. The areas covered from green to red represent the distribution of fixation time from short to long. The small bule circles are gaze points, and the yellow lines between the gaze points represent saccades. This visual information offers a rough judgment of the participant's visual searching patterns, such as the time allocation and the number of toggles between the matrix and response options areas. The constructive matching strategy selection probability of three participants on the three items is also presented above the diagrams. Taking Participant 1 as an example, the eye-tracking diagrams revealed that for items 1 and 36, they spent much more time in the matrix area than in the response options area, and we can roughly infer

[9] The data presented in Fig. 6 were obtained by computing averages, and there may be some information loss. Additional smooth 3D surface plots and a corresponding scatter plot of all information were presented in Figure E5 in Appendix E.

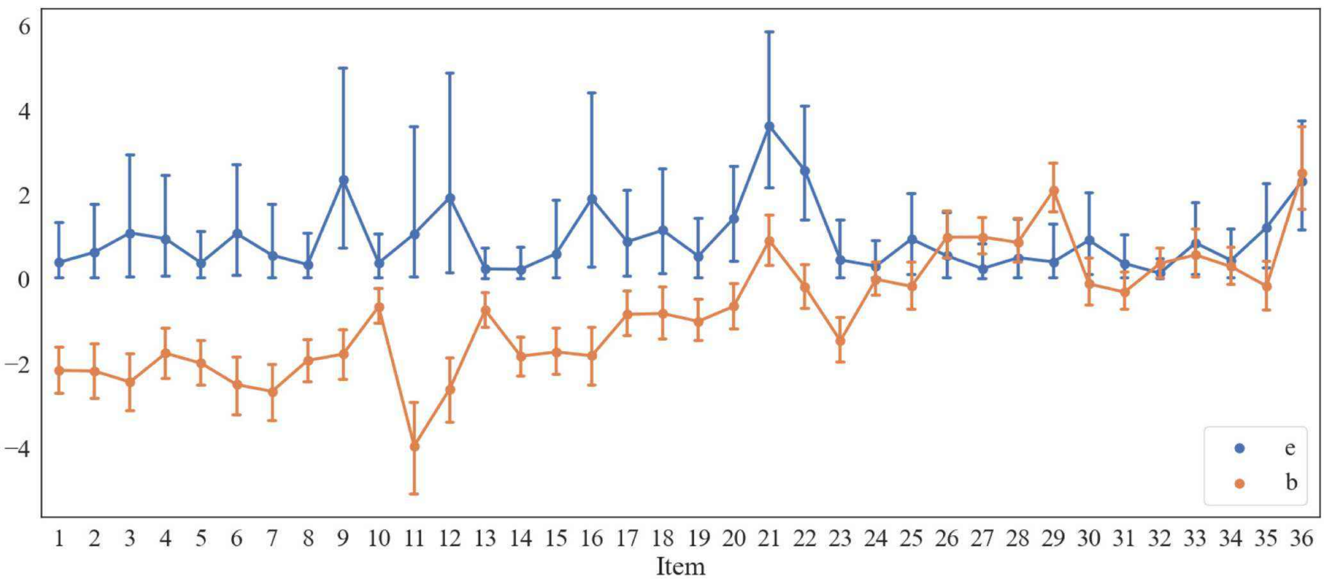

Fig. 4. Posterior mean and 95% highest posterior density of item difficulty and strategy sensitivity parameters of 36 Items.
Note. b = item difficulty; e = item strategy sensitivity.

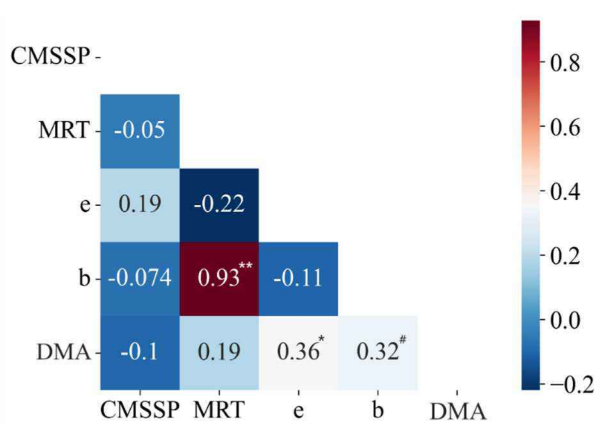

Fig. 5. Spearman rank correlation among different indicators on the item level.
Note. CMSSP: the mean probability of participants using the constructive matching strategy per item; MRT: the mean response time across participants per item; DMA: the difference between the mean abilities of two strategy groups per item; e: strategy sensitivity parameter; b: item difficulty. The number of elements in each indicator was 36. **: $p < 0.01$; *: $p < 0.05$; #: $p = 0.055$ (marginal significant).

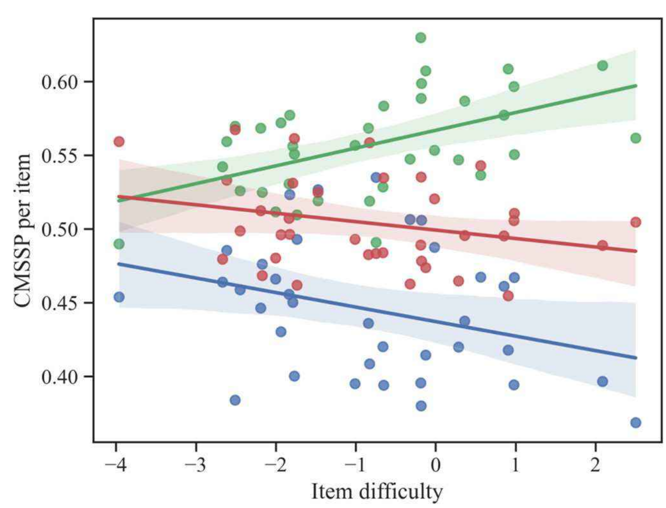

Fig. 6. Scatterplot of item difficulty and CMSSP for three intelligence level groups.
Note. CMSSP: constructive matching strategy selection probability per item. The green, red, and blue represent high-, medium-, and low-intelligence groups, respectively.

Table 2
Regression analyses predicting constructive matching strategy selection probability by intelligence, item difficulty, and their interaction.

Independent variable	*coef*	*t*	*p*
Intercept	0.499	107.531	0.000
θ_i	0.055	9.175	0.000
b_j	−0.002	−0.682	0.495
$\theta_i \times b_j$	0.011	2.906	0.004

Note. F-statistic = 29.36; Prob(F-statistic) = 7.44e^{-19}; *N = 6912.*

that they used the constructive matching strategy on these two items. Such an inference matched the high selection probability of the constructive matching strategy estimated by the MEM. By contrast, the small difference in time spent by Participant 1 in the matrix and response options areas of item 34 and the high number of toggles between the two areas seemed to indicate that they used the response elimination strategy on this item, which also matched the low selection probability of the constructive matching strategy estimated by the MEM.

Of course, not all participants' strategy selections were as clear as those of Participant 1. For example, the estimated probability of Participant 138 choosing the constructive matching strategy on item 1 was 0.45; their eye-tracking diagram also showed that they spent some time both in the matrix and response options areas and had a certain number of toggles between the two areas, which led us to draw no easy conclusions about which one of the two strategies they actually used. One possible reason for this is that this participant tried to use two strategies to respond to this item.

Furthermore, for each item, to distinguish between participants with clear and ambiguous strategy use, we divided them into three strategy groups according to whether the estimated value of $P(m_{ij})$ was significantly different from 0.5: those significantly >0.5 were the constructive matching strategy group, those significantly <0.5 were the response

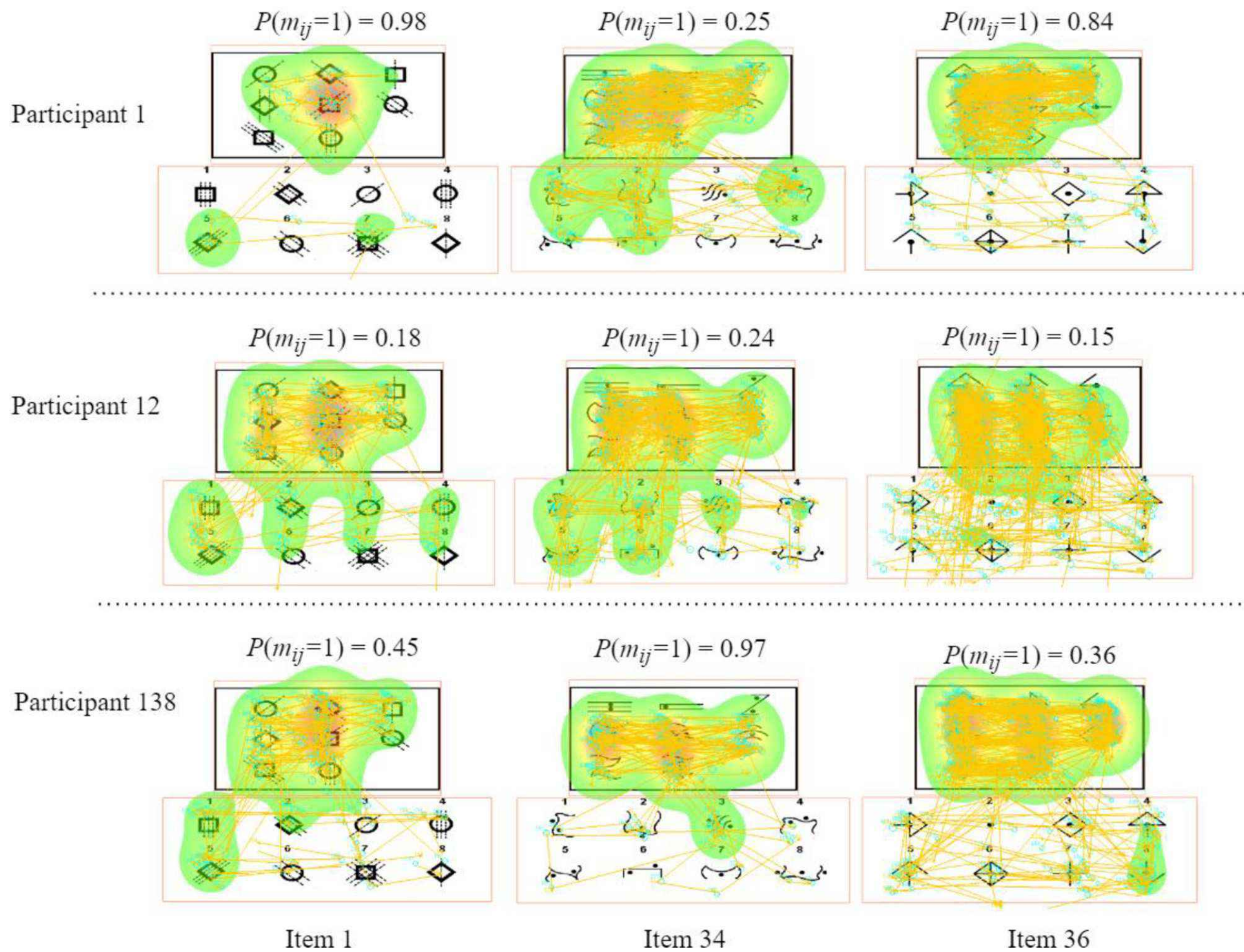

Fig. 7. Eye-tracking diagrams (heat maps and gaze plots) of three participants on three items and corresponding constructive matching strategy selection probabilities.
Note. $P(m = 1)$: constructive matching strategy selection probability. The area covered from green to red represents the distribution of fixation time, the small blue circles are gaze points, and the line between the gaze points represents the saccade. The diagram (heat map and gaze plot) was automatically generated using EyeLink Data Viewer software based on the recorded eye-tracking data.

elimination strategy group, and those not significantly different from 0.5 were the ambiguous strategy use group. Fig. 8 displays the sample sizes of the three strategy groups on each item; the two groups with clear strategy use accounted for approximately 78% of the total, indicating the majority of participants' strategy use can be identified.

Fig. 9 further displays the mean differences in intelligence, response

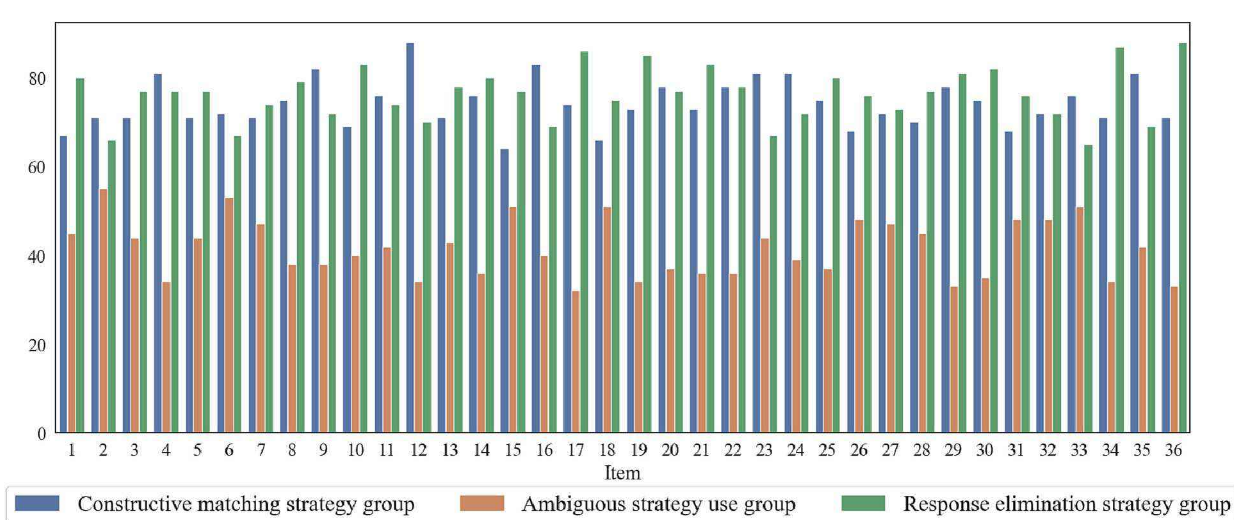

Fig. 8. Sample sizes of three strategy groups on each item.

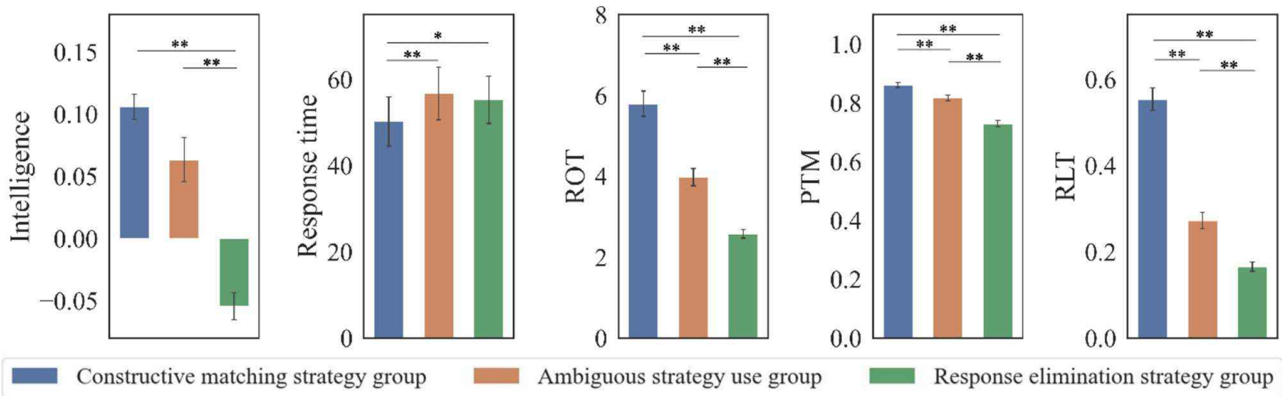

Fig. 9. Bar charts of three groups on intelligence, response time, ROT, PTM, and RLT.
Note. ROT: rate of toggling; PTM: proportional time on matrix; RLT: rate of latency to first toggle; error bar: 95% confidence interval; **: $p < 0.01$; *: $p < 0.05$.

time, and three eye-tracking measures for the three strategy groups of participants across all items.[10] First, the results indicated that the mean intelligence and all three-mean eye-tracking measures across participants decreased in descending order in the constructive matching strategy group, the ambiguous strategy use group, and the response elimination strategy group. By contrast, the mean response time across participants for the constructive matching strategy group was the lowest and less than the similar mean response times for the response elimination strategy and ambiguous strategy use groups. Further, the results of the one-way repeated measures ANOVA[11] showed a significant main effect of the group on the five indicators in Fig. 9 (intelligence: F (2, 70) = 26.65, $p < 0.01$, $\eta_p^2 = 0.43$; response time: F (2, 70) = 9.12, $p < 0.01$, $\eta_p^2 = 0.21$; ROT: F (1.51, 52.8) = 145.09, $p < 0.01$, $\eta_p^2 = 0.81$; PTM: F (1.54, 53.95) = 923.38, $p < 0.01$, $\eta_p^2 = 0.96$; RLT: F (1.31, 45.87) =501.41, $p < 0.01$, $\eta_p^2 = 0.94$), indicating significant differences in these indicators among the three strategy groups. Further, Bonferroni post hoc comparisons showed a significant difference between any two of the three strategy groups on each mean eye-tracking measure (all $\hat{p}$ s < 0.01[12]). For the mean intelligence, the differences between the constructive matching strategy group and the ambiguous strategy use group were not significant ($\hat{p} = 0.218$), but both were significantly higher than the response elimination strategy group ($\hat{p}$ s < 0.01). In addition, the mean response time of constructive matching strategy group was significantly less than that of ambiguous strategy use group ($\hat{p} < 0.01$) and that of response elimination strategy group ($\hat{p} = 0.017$); however, there was no significant difference in the mean response time between ambiguous strategy use group and elimination strategy group ($\hat{p} = 1$). Overall, primarily focusing on the comparison between the two groups with clear use of strategies, it can be found that compared with the participants using the response elimination strategy, the participants using the constructive matching strategy had a higher mean intelligence level, a lower mean response time, a higher mean ROT, a higher mean PTM, and a higher mean RLT. Furthermore, considering that each of the three eye-tracking measures of the ambiguous strategy use group was between the constructive matching strategy group and the response elimination strategy group, it can be inferred that participants in the ambiguous strategy use group might use a combination of the constructive matching strategy and the response elimination strategy.

Finally, the two staff members endorsed 88% and 86% of the strategy identification results from the MEM, respectively, and both endorsed 81% of the strategy identification results (see Table D1 in Appendix D). This provided some evidence of the validity of the strategy identification results of the proposed model.

3.2. Additional results

In the above analysis, we found that the ROT, which has been studied as a concern in previous studies (e.g., Laurence et al., 2018), has no significant effect on the choice of strategy. To further understand this result, it might be important to track what response option one toggles to, such as distinguishing between correct and incorrect options (e.g., Kroczek, Ciechanowska, & Chuderski, 2022). Currently, the ROT confounds the toggle to the correct and incorrect options, and perhaps because of this fact, the ROT appeared non-significant in the choice of strategy in the MEM. To this end, we conducted additional supplementary analyses by dividing the ROT into two sub-measures: the rate of toggling between the matrix area and the correct option ($ROTCO = T^{item} / GCO$, where GCO is the number of toggles between the matrix area and the correct option) and the rate of toggling between the matrix area and the incorrect options ($ROTIO = T^{item} / GIO$, where GIO is the number of toggles between the matrix area and the incorrect options). The former reflects how long it takes to toggle once between the matrix area and the correct option, and the latter reflects that between the matrix area and the incorrect options. We first explored the relationship between these two sub-measures and strategy selection probability and then incorporated them into the MEM to explore their predictions of the strategy selection probability.

Fig. E6 in Appendix E presents the Pearson correlation among the CMSSP, ROTCO, and ROTIO. The ROTIO had a significant positive correlation with CMSSP, while the ROTCO had a non-significant correlation with CMSSP. Such results suggested that the ROTIO appears to help reflect participants' strategy choices. Fig. E7 in Appendix E further presents the differences in mean values of ROTCO and ROTIO among the three strategy groups. One-way repeated measures ANOVA showed a significant main effect of the group (ROTCO: F (2, 70) = 4.45, $p < 0.05$, $\eta_p^2 = 0.11$; ROTIO: F (1.56, 54.73) = 130.5, $p < 0.01$, $\eta_p^2 = 0.79$). Further, Bonferroni post hoc comparisons presented a significant difference between any two of the three strategy groups on ROTIO (all $\hat{p}$ s < 0.01). For the ROTCO, only the difference between the ambiguous strategy use group and the response elimination strategy group was significant ($\hat{p} < 0.05$).

Table E3 in Appendix E presents the estimated weight parameters and two relative model-data fit indices for the three models, including two models (i.e., MEM-a and MEM-b) containing three eye-tracking measures (PTM, RLT, and one of the two sub-measures) and one

[10] For each variable, we calculated the average of the three groups of participants on each item to make the data of the three groups on each variable paired (i.e., for each variable, the data length for each group is 36).
[11] One-way repeated measures ANOVA was performed using the "rsratix" package in R. Spherical correction was performed automatically using "get_ anova_table()" to make the variance homogeneous (Kassambara, 2020; details see https://rpackgs.datanovia.com/rstatix).
[12] $\hat{p}$ is the p-value after Bonferroni adjustment.

model (i.e., MEM-ab) containing four eye-tracking measures (PTM, RLT, and both two sub-measures). The results showed that the ROTCO and ROTIO played opposite roles in predicting the selection probability of constructive matching strategy: ROTCO is a positive prediction, while ROTIO is a negative prediction. Such results indicated that participants were more likely to use the constructive matching strategy when they consulted the incorrect options less often per unit of time or the correct option more often per unit of time. In addition, when splitting the ROT into ROTIO and ROTCO, the model-data fit increased. In particular, when the ROTIO was incorporated, the predictive effect of the PTM and RLT on the strategy selection probability became non-significant; such results may imply that the information provided by ROTIO overlaps with the information provided by the PTM and RLT. Furthermore, as shown in Table E4 in Appendix E, there is relatively low consistency of strategy identification for models that include the ROTIO versus those that do not, indicating that the ROTIO may provide additional information different from that provided by the PTM and RLT.

To further investigate the strategy identification accuracy of the model with the ROT split (i.e., the MEM-ab) and the model without the ROT split (i.e., the MEM), we randomly selected some participants' eye-tracking diagrams and found that the latter's identification results for strategies matched the patterns presented in the eye-tracking diagrams more than the former's identification results for strategies. Fig. E8 in Appendix E displays the eye-tracking diagrams of four participants on four items (the correct responses to the item in Figures (a), (b), (c), and (d) are options 5, 1, 6, and 1, respectively), and the strategy identification results of the two models in these four diagrams were different. For example, for Fig. E8(a), the visual search pattern shows that this participant spent more time in the matrix area than the options area and had a few toggles between the two areas. Thus, this visual search pattern is more consistent with the constructive matching strategy (i.e., the identification result of MEM) than the response elimination strategy. However, perhaps because this participant toggled between the matrix area and the correct option very few times led to a large ROTCO, which led to an identification by the MEM-ab as a response elimination strategy. In addition, for Fig. E8(b), the visual search pattern of it displays that this participant spent more time in the matrix area than the options area but had many toggles between the two areas, indicating this participant tried to find the rule in the matrix area and also tried to compare between the options. Hence, this visual search pattern tended to conform to the constructive matching strategy while being ambiguous, which is consistent with the identification result of MEM. However, perhaps because this participant toggled between the matrix area and the incorrect options very few times and correct option many times lad to a large ROTIC and a small ROTCO, which further led to a fairly certain identification of the MEM-ab as a constructive matching strategy.

Furthermore, the visual search pattern of Fig. E8(c) shows that the participant used a constructive matching strategy but induced a wrong rule from the matrix area. However, the MEM-ab identified it as a response elimination strategy because this participant had multiple toggles between the matrix area and the objectively incorrect but subjectively correct option. Finally, for Fig. E8(d), the identification results of the two models were also opposite, most likely also because the MEM-ab focused more on there is a lot toggles between the matrix area and the correct option. In summary, the strategy identification results of the MEM-ab would be heavily influenced by the frequency with which participants toggle to the incorrect or correct options from the matrix area, ignoring the difference between the elapsed time of the two interest areas.

4. Summary and discussion

4.1. Summary

The APM is a valid measurement instrument of intelligence, and previous studies have investigated the role of cognitive strategies in responding to these items. This study proposes a multi-strategy psychometric model incorporating eye-tracking measures (i.e., PTM, ROT, and RLT). By jointly analyzing item responses and eye-tracking measures, the proposed model can measure each participant's intelligence and identify the cognitive strategy used by each participant on each item in the APM at the same time. The proposed model follows the theory-driven modeling logic and provides a new way to study cognitive strategy in the APM by presenting objective and quantitative results based on existing findings on the correspondence between cognitive strategies and eye-tracking measures. Theoretically, compared with eye-tracking data mining approaches, the proposed model is more interpretable and has fewer data volume requirements, which makes it more suitable for small-scale psychological experimental studies. Compared with self-reporting approaches, the proposed model can more objectively identify participants' cognitive strategies for each item.

The findings of this study can be summarized as follows. First, the MEM fitted the data better than the Rasch model while ensuring that the same latent trait (i.e., intelligence) was measured, indicating that the participants do use different cognitive strategies in responding to items in the APM. Second, the effects of PTM and RLT on the constructive matching strategy selection probability were positive and higher for the former than the latter, while the effect of ROT was negligible. Third, the average intelligence of participants using the constructive matching strategy was higher than that of participants using the response elimination strategy. Fourth, participants with higher intelligence were more likely to use the constructive matching strategy. Fifth, the choice of strategy has different magnitudes of effect on different items in the APM; namely, choosing to respond to items with the constructive matching strategy did not result in significant gains on every item in the APM.

Sixth, for all participants, there was no significant correlation between item difficulty and the probability of participants choosing the constructive matching strategy. However, there was a significant interaction effect between item difficulty and intelligence. High-intelligence participants increased their use of the constructive matching strategy as item difficulty increased, whereas low-intelligence participants tended to decrease their use of the constructive matching strategy as item difficulty increased; in addition, participants in the medium-intelligence group had a similar trend to participants in the low-intelligence group, but the trend was not significant.

Seventh, participants took significantly less time to use the constructive matching strategy than to use the response elimination strategy. Eighth, the use of the two cognitive strategies coexisted in responding to the items in the APM; some participants were more likely to use one of them, while some participants might try both strategies. Ninth, after splitting the ROT into two sub-measures at the option level (i.e., ROTCO for the correct option and ROTIO for the incorrect options), the two predict the constructive matching strategy selection probability in opposite ways: the smaller the former or the larger the latter, the greater the probability. Meanwhile, there are some differences in the strategy identification results between the model with ROT disassembled (i.e., the MEM-ab) and the model without ROT disassembled (i.e., the MEM). Overall, the strategy identification results of the latter match the visual search pattern of the eye-tracking diagrams more than those of the former. Finally, the validity evidence based on manual judgment also indicated that the proposed model was, to some extent, effective for identifying cognitive strategies in the APM.

4.2. Discussion

The findings of this study show consistency as well as disagreement with previous studies. For example, consistent with the findings of Bethell-Fox et al. (1984) and Vigneau et al. (2006), the use of the constructive matching strategy has a significant positive correlation with participants' intelligences, which means high intelligence participants tended to use the constructive matching strategy. Consistent with the findings of Vigneau et al. (2006), for all participants, the usage of strategy was independent from item difficulty; however, further analysis of this study showed that there was an interaction between difficulty and intelligence, with participants of different intelligence levels choosing different strategies when treating difficult items. For low-intelligence participants, our findings support the findings of Jarosz et al. (2019) and Gonthier and Roulin (2020) that participants decrease their use of the constructive matching strategy as item difficulty increases. In this regard, Bethell-Fox et al. (1984) explained that when responding to difficulty items, low-intelligence participants tend to shift from using the constructive matching strategy to using the response elimination strategy, which may be due to the failure to induce the rules from the matrix area.

The findings of this study were found in untimed APM, which may lead to some time-related findings that are different from those of existing studies. For example, contrary to the findings of Gonthier and Roulin (2020) and Gonthier and Thomassin (2015), the results of this study showed that participants who used the constructive matching strategy spent less time than those who used the response elimination strategy. One of the possible reasons for this discrepancy is that in their studies, participants were asked to complete the APM within a specific time limit, whereas, in this study, there was no time limit. Some previous studies have pointed out that time pressure can affect strategy selection, causing individuals to sacrifice response accuracy in order to make a quick decision in decision-making and cognitive tasks, namely, the speed-accuracy trade-off (e.g., Caviola, Carey, Mammarella, & Szucs, 2017; Starcke & Brand, 2012; Szalma, Hancock, & Quinn, 2008). Furthermore, Fehrenbacher and Smith (2014) found that time pressure reduced participants' attention to available information, reducing the retrieval of useful information. Hence, when participants perceive time pressure in the timed APM, they may reduce their time spent on the matrix. Instead, they use the response elimination strategy to go for a seemingly correct answer and make a quick choice, especially when they cannot identify the rules in the matrix. This may result in less time for the response elimination strategy than the constructive matching strategy.

In contrast to the findings of some previous studies (e.g., Laurence et al., 2018; Vigneau et al., 2006) that suggested that ROT was one of the best measures for distinguishing strategy and predicting test outcomes (i.e., raw scores), our findings indicate that ROT has no significant effect on predicting strategy selection when PTM and RLT are present. One of the possible reasons for this discrepancy is that most existing studies did not directly explore the relationship between these three eye-tracking measures and the strategy used, as the MEM did in this study, but instead used indirect speculation to obtain the conclusion. For example, Laurence et al. (2018) explored the predictive role of several indices, including the ROT, on raw scores and found that the ROT was the best predictor of raw scores. They then inferred that the ROT was the best predictor of strategy use by combining the characteristics of two cognitive strategies from existing studies (e.g., Vigneau et al., 2006). Another reason may be that the ROT used in current study confounds the toggle to the correct and incorrect options, resulting the predictive effects of the two option-level ROT sub-measures (i.e., ROTCO and ROTIO) on the strategy selection probability to cancel each other out. Although the supplementary analysis supports this view, it also leads to some differences in the strategy identification results for the two models before and after the ROT split. Given that the model using option-level ROT may be biased in strategy identification (e.g., participants selected the wrong answer using the constructive matching strategy would lead to a large GIO and a small ROTIO, leading the MEM-ab biasedly identify the participants as using a response elimination strategy) and that little study has focused on the role of option-level ROT in strategy identification, the effectiveness of option-level ROT in strategy identification is yet to be further investigated in the future.

In addition to the two constructive strategies investigated in this study, Jarosz et al. (2019) pointed out that participants may also use another cognitive strategy in the APM—the isolate-and-eliminate strategy. The isolate-and-eliminate strategy can be treated as a combination of the constructive matching strategy and the response elimination strategy, in which "participants would establish the rules governing one feature of items in the problem matrix, use that rule to eliminate potential responses from the response bank, and then proceed with another feature, until only one item was left" (Jarosz et al., 2019, p. 5). In this study, we found that the estimates of the probability of strategy selection of some participants (i.e., $P(m_{ij})$) were within the ambiguous region of about 0.5 (i.e., no significant difference from 0.5). Considering that each of the three eye-tracking measures of these participants was between those using constructive matching strategy and those using response elimination strategy, it can be inferred that these participants may have tried to use both strategies simultaneously (e.g., strategy shift occurs during problem-solving). Such an interpretation may provide a new perspective for understanding the isolate-and-eliminate strategy.

Despite the promising results of this study, there are some limitations of this study that need to be addressed in further studies. First, consistent with most studies on cognitive strategy in the APM, this study focused on only two cognitive strategies—the constructive matching and response elimination strategies. However, in practice, more types of cognitive strategies may exist when responding to items on the APM, such as the isolate-and-eliminate strategy (Jarosz et al., 2019) and the goal management strategy (Carpenter et al., 1990). Carpenter et al. (1990) believed that the response of the APM involves a goal management strategy, which means breaking down the overall goal into sub-goals, and is demonstrated in the APM test by participants inducing one rule at a time. Alternatively, DeShon, Chan, and Weissbein (1995) suggested that there were two other strategies in APM, one relying on visuospatial processes and the other relying on verbal-analytic processes. Theoretically, the MEM proposed in this study is not limited to two strategies; thus, the performance of the MEM with more than two strategies is worthy of further exploration.

Second, the relationship between intelligence and the strategy use (or other indices) obtained in this study is based on a single batch of data obtained from participants solving APM items. Consequently, the latent trait estimates were indirectly inferred from the same batch data from which the strategy identifies. Although the MEM assumes that participants' choice of cognitive strategy is not theoretically related to their intelligence but is reflected by behavioral indicators of the problem-solving process (i.e., eye-tracking measures), because of their shared source of data, some unknown dependencies between ability and strategy might have emerged. Since we did not obtain more generalized intelligence of the participants through multiple external measures, it remains to be further verified whether this study's findings are robust and can be generalized to a broader domain.

Third, the three composite eye-tracking measures used in this study are based on the participants' complete responses to each item, ignoring changes in cognitive states during problem-solving. Therefore, the proposed model cannot be used to explore the possible strategy shifts in problem-solving (e.g., the proposed model cannot explore how many participants use the response elimination strategy after a failed attempt to use the constructive matching strategy). Meanwhile, the proposed model may not accurately distinguish between different cognitive processes because similar values of these eye-tracking measures do not imply similar cognitive processes. In the future, eye-tracking data's high temporal accuracy nature can be fully leveraged using time-series eye-tracking measures instead of the composite eye-tracking measures to achieve dynamic identification of strategy usage.

Finally, participants were randomly selected from the same university. Although the sample size in this study was larger than in other related studies, the results may not be universal due to the homogeneity among participants.

Overall, following the theory-driven modeling logic, this study proposed a multi-strategy psychometric model incorporating three eye-tracking measures (i.e., PTM, ROT, and RLT) to measure each participant's intelligence and identify the cognitive strategy used by each participant on each item in the APM at the same time. Several main findings were revealed from an eye-tracking-based APM study using the proposed model: (1) The effects of PTM and RLT on the constructive matching strategy selection probability were positive and higher for the former than the latter, while the effect of ROT was negligible. (2) The average intelligence of participants who used the constructive matching strategy was higher than that of participants who used the response elimination strategy, and participants with higher intelligence were more likely to use the constructive matching strategy. (3) High-intelligence participants increased their use of the constructive matching strategy as item difficulty increased, whereas low-intelligence participants decreased their use as item difficulty increased. (4) Participants took significantly less time using the constructive matching strategy than the response elimination strategy.

Funding

This work was supported by the National Natural Science Foundation of China [grant numbers 31900795].

Declaration of Competing Interest

None.

Data availability

I have shared the link to my data in the manuscript.

Appendix A. Supplementary data

Supplementary materials to this article can be found online at https://doi.org/10.1016/j.intell.2023.101782.

References

Arendasy, M. E., & Sommer, M. (2013). Reducing response elimination strategies enhances the construct validity of figural matrices. *Intelligence, 41*(4), 234–243.
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.
Beuchert, A. K., & Mendoza, J. L. (1979). A Monte Carlo comparison of ten item discrimination indices. *Journal of Educational Measurement*, 109–117.
Cameron, L. D., & Jago, L. (2013). Cognitive strategies. In M. D. Gellman, & J. R. Turner (Eds.), *Encyclopedia of behavioral medicine*. New York, NY: Springer.
Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven progressive matrices test. *Psychological Review, 97*, 404–431.
Caviola, S., Carey, E., Mammarella, I. C., & Szucs, D. (2017). Stress, time pressure, strategy selection and math anxiety in mathematics: A review of the literature. *Frontiers in Psychology, 8*, 1488.
Chiu, I., & Shu, L. H. (2010, January). Potential limitations of verbal protocols in design experiments. In *, vol. 44137. International design engineering technical conferences and computers and information in engineering conference* (pp. 287–296).
Cook, L. K., & Mayer, R. E. (1983). Reading strategies training for meaningful learning from prose. *Cognitive Strategy Research: Educational Applications*, 87–131.
von Davier, M. (2010). Mixture distribution item response theory, latent class analysis, and diagnostic mixture models. In S. E. Embretson (Ed.), *Measuring psychological constructs: Advances in model-based approaches* (pp. 11–34). Washington, DC: American Psychological Association.
Dehn, M. J. (2017). How working memory enables fluid reasoning. *Applied Neuropsychology: Child, 6*(3), 245–247.
DeShon, R. P., Chan, D., & Weissbein, D. A. (1995). Verbal overshadowing effects on Raven's advanced progressive matrices: Evidence for multidimensional performance determinants. *Intelligence, 21*(2), 135–155.
Embretson, S. E., & Reise, S. P. (2013). *Item response theory*. Psychology Press.
Engelhart, M. D. (1965). A comparison of several item discrimination indices. *Educational and Psychological Measurement, 2*(1), 69–76.
Fehrenbacher, D., & Smith, S. (2014). Behavioural affect and cognitive effects of time-pressure and justification requirement in software acquisition: Evidence from an eye-tracking experiment. In *Proceedings of the AMCIS 2014, Savannah*.
Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis*. CRC Press.
Gonthier, C. (2022). Cross-cultural differences in visuo-spatial processing and the culture-fairness of visuo-spatial intelligence tests: An integrative review and a model for matrices tasks. *Cognitive Research: Principles and Implications, 7*(1), 11.
Gonthier, C., & Roulin, J.-L. (2020). Individual 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.
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.
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.
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.
Hessels, M. G. P., Vanderlinden, K., & Rojas, H. (2011). Training effects in dynamic assessment: A pilot study as eye movement as indicator of problem solving behavior before and after training. *Educational and Child Psychology, 28*(2), 101–113.
Jarosz, A. F., Raden, M. J., & Wiley, J. (2019). Working memory capacity and strategy use on the RAPM. *Intelligence, 77*, Article 101387.
Jarosz, A. F., & Wiley, J. (2012). Why does working memory capacity predict RAPM performance? A possible role of distraction. *Intelligence, 40*(5), 427–438.
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.
Kane, M. J., Hambrick, D. Z., & Conway, A. R. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin, 131*, 66–71.
Kassambara, A. (2020). *Package 'rstatix'. R topics documented*.
Kpolovie, P. J., & Emekene, C. O. (2016). Item response theory validation of advanced progressive matrices in Nigeria. *British Journal of Psychology Research, 4*(1), 1–32.
Kroczek, B., Ciechanowska, I., & Chuderski, A. (2022). Uncovering the course of analogical mapping using eye tracking. *Cognition, 225*, Article 105140.
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*.
de La Torre, J., & Douglas, J. A. (2008). Model evaluation and multiple strategies in cognitive diagnosis: An analysis of fraction subtraction data. *Psychometrika, 73*(4), 595–624.
Laurence, P. G. (2021). *Individual differences in cognitive abilities affects strategies revealed by eye movements in tasks with answer choices* [Unpublished doctoral dissertation]. University of Mackenzie Presbyterian
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.
Lemaire, P., & Reder, L. (1999). What affects strategy selection in arithmetic? The example of parity and five effects on product verification. *Memory & Cognition, 27*(2), 364–382.
Li, C., Ren, X., Schweizer, K., & Wang, T. (2022). Strategy use moderates the relation between working memory capacity and fluid intelligence: A combined approach. *Intelligence, 91*, Article 101627.
Loesche, P. (2020). Raven's advanced progressive matrices. In V. Zeigler-Hill, & T. K. Shackelford (Eds.), *Encyclopedia of personality and individual differences*. Cham: Springer.
Lovett, M. C., & Anderson, R. (1995). Making heads or tails out of selecting problem-solving strategies. In *Proceedings of the seventeenth annual conference of the cognitive science society* (pp. 265–270). Hillsdale, NJ: Lawrence Erlbaum Associates.
Ma, W., & Guo, W. (2019). Cognitive diagnosis models for multiple strategies. *The British Journal of Mathematical and Statistical Psychology, 72*(2), 370–392.
Man, K., & Harring, J. R. (2019). Negative binomial models for visual fixation counts on test items. *Educational and Psychological Measurement, 79*(4), 617–635.
Marshalek, B., Lohman, D. F., & Snow, R. E. (1983). The complexity continuum in the radex and hierarchical models of intelligence. *Intelligence, 7*, 107–127.
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), 1–10.
Mislevy, R. J., & Verhelst, N. (1990). Modeling item responses when different subjects employ different solution strategies. *Psychometrika, 55*, 195–215.
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.
Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Nielson & Lydiche.
Raven, J., Raven, J. C., & Court, J. (1998). *Manual for Raven's progressive matrices and vocabulary scales*. Oxford, UK: Oxford Psychologists Press.
Starcke, K., & Brand, M. (2012). Decision making under stress: A selective review. *Neuroscience & Biobehavioral Reviews, 36*(4), 1228–1248.
Szalma, J. A., Hancock, P. A., & Quinn, S. (2008, September). A meta-analysis of the effect of time pressure on human performance. In *, vol. 52. Proceedings of the human factors and ergonomics society annual meeting* (pp. 1513–1516). Sage CA: Los Angeles, CA: SAGE Publications. No. 19.
Thibaut, J. P., & French, R. M. (2016). Analogical reasoning, control and executive functions: A developmental investigation with eye-tracking. *Cognitive Development, 38*, 10–26.
Vehtari, A., Gelman, A., & Gabry, J. (2016). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing, 27*(5), 1413–1432.

Vigneau, F., Caissie, A. F., & Bors, D. A. (2006). Eye-movement analysis demonstrates strategic influences on intelligence. *Intelligence, 34*(3), 261–272.

Waschl, N. A., Nettelbeck, T., Jackson, S. A., & Burns, N. R. (2016). Dimensionality of the Raven's advanced progressive matrices: Sex differences and visuospatial ability. *Personality and Individual Differences, 100*, 157–166.

Wedel, M., Pieters, R., & van der Lans, R. (2022). Modeling eye movements during decision making: A review. *Psychometrika*, 1–33.

Zhan, P., Jiao, H., Man, K., & Wang, L. (2019). Using JAGS for Bayesian cognitive diagnosis modeling: A tutorial. *Journal of Educational and Behavioral Statistics, 44*(4), 473–503.

Zhan, P., Man, K., Wind, S. A., & Malone, J. (2022). Cognitive diagnosis modeling incorporating response times and fixation counts: Providing comprehensive feedback and accurate diagnosis. *Journal of Educational and Behavioral Statistics, 47*(6), 736–776.