Journal of Experimental Psychology: General

ISSN: 0096-3445

2024, Vol. 153, No. 3, 689–705
https://doi.org/10.1037/xge0001518

Development of Reasoning Performance in Raven's Matrices Is Grounded in the Development of Effective Strategy Use

Corentin Gonthier[1, 2], Kahina Harma[3], and Zdenka Gavornikova-Baligand[4]
[1] Department of Psychology, Laboratoire de Psychologie des Pays de la Loire (LPPL UR 4638), Chemin de la Censive du Tertre, Nantes Université
[2] Institut Universitaire de France
[3] Centre de Recherche sur l'Education, les Apprentissages et la Didactique (CREAD UR 3875), University of Brest
[4] Laboratoire de Psychologie: Cognition, Comportement, Communication (LP3C UR 1285), University of Brest

Performance in reasoning tasks such as Raven's matrices experiences a dramatic increase over cognitive development, but the mechanisms responsible for this increase are unknown. Many cognitive processes are involved in a matrix task and could potentially change with age; strategy use appears to be a good candidate, as it typically improves over development and has a large impact on reasoning performance in adults. The present study tested the role of effective strategy use in Raven's standard progressive matrices in groups of 6-, 8-, 10-, 12-, 14-, 16-, and 18-year-olds (total $N = 474$). Strategy use was assessed with behavioral measures of gaze patterns in Raven's matrices. We also measured working memory capacity (WMC), a good predictor of strategy use in adults, using a battery of complex spans. The results showed that the effective strategy of constructive matching substantially increased with age, along with performance. Strategy use mediated over half the effect of age on reasoning performance. Older participants were also better at adapting strategy use to difficulty of the problems. Effective strategy use was beneficial to the same extent for participants of all ages. Age-related improvements in strategy use occurred in tandem with improvements in WMC, but did not appear to be primarily driven by them. Overall, our results indicate that strategy use is a critical underpinning of reasoning performance in children as well as in adults, and that theories of cognitive development of reasoning have to consider the central role of strategy use.

Public Significance Statement

Raven's matrices and similar matrix reasoning tasks are the test's most representative of fluid intelligence, and are used in the majority of both intelligence research and clinical assessments of intelligence (either alone or as part of a larger battery, such as Wechsler's scale of intelligence). Understanding the origin of individual differences and age-related differences of performance is therefore critical to correctly interpreting differences of reasoning performance between individuals or groups in both research and applied settings.

Keywords: fluid intelligence, Raven's matrices, strategy use, working memory capacity, cognitive development

Fluid reasoning performance is an excellent predictor of real-life outcomes in children (such as academic achievement: Laidra et al., 2007; learning: Primi et al., 2010; or even violent behaviors and drug use: Huepe et al., 2011). Critically, performance on fluid reasoning tasks also increases to a considerable extent over the course of cognitive development. The finding of a developmental increase of intelligence is as old as developmental research (Binet & Simon, 1904), but probably the easiest way to illustrate this point is to examine the norms of fluid intelligence tests. Let us take the example of Raven's matrices, the test most representative of fluid intelligence

This article was published Online First December 7, 2023.
Corentin Gonthier https://orcid.org/0000-0001-8573-0413
The data appearing in this article were presented at the 24th Journées Internationales de Psychologie Différentielle Conference (Aix-en-Provence, June 2022). This study was partly funded by Grant 18C0354 from the city of Rennes (*Allocation d'Installation Scientifique*, Rennes Métropole, France). The authors warmly thank Mélina Le Guellec and Thibault Bernier who contributed to data collection; all the schools who welcomed the testing sessions; and all the children and parents who agreed to participate in the study. All the data for this study are available on the Open Science Framework at https://osf.io/zyemb/ (Gonthier, 2022; available upon publication).

Corentin Gonthier served as lead for conceptualization, data curation, formal analysis, funding acquisition, methodology, project administration, resources, software, supervision, visualization, and writing–original draft. Kahina Harma served in a supporting role for writing–review and editing. Zdenka Gavornikova-Baligand served in a supporting role for writing–review and editing. Kahina Harma and Zdenka Gavornikova-Baligand contributed equally to investigation.
Correspondence concerning this article should be addressed to Corentin Gonthier, Department of Psychology, Laboratoire de Psychologie des Pays de la Loire (LPPL UR 4638), Chemin de la Censive du Tertre, Nantes Université, BP 81227, 44312 Nantes Cedex 3, France. Email: corentin.gonthier@univ-nantes.fr

(Carpenter et al., 1990) and one with a large amount of available developmental data. In the British 1979 standardization sample of Raven's standard progressive matrices (SPM; J. Raven et al., 1998), median scores increased from 15 out of 60 correct answers at 6 years old, to 45 out of 60 correct answers at 14 years old ($N = 3{,}250$ children; J. Raven, 2000, Figure 3). Similar increases with age have been found in other norming datasets since the very first uses of Raven's matrices (e.g., J. C. Raven, 1941), as well as in developmental research (Perret & Dauvier, 2018).

What leads reasoning performance in Raven's matrices to triple over just 8 years or cognitive development? Identifying the underpinnings of such a dramatic increase in fluid reasoning performance is of major interest, both to the understanding of cognitive development, and to the understanding of cognitive processes underlying reasoning. It would hardly be satisfying to simply argue that children "get more intelligent" during cognitive development; besides, improvement in performance does not necessarily reflect improvement in intelligence (Hayes et al., 2015). Instead, it is much more informative to consider which of the various determinants of performance in Raven's matrices drive this developmental change. This question can be approached in terms of the quantitative abilities involved in the task, such as processing speed (e.g., Kail, 2007), working memory (e.g., de Ribaupierre & Lecerf, 2006), or inhibition (e.g., Perret, 2015); or in terms of the mechanisms qualitatively involved in the process of solving an item (e.g., Gonthier & Roulin, 2020; Styles, 2008). The present study focuses on the latter possibility, with the hypothesis that gradual improvements in performance are accompanied by qualitative changes of behavior in the task, as reflected in a progressive increase in effective strategy use.

Strategy Use in Matrix Reasoning Tasks

A strategy is the procedure used by a subject to reach a high-level goal in a task (Lemaire & Reder, 1999), consciously or not, over and above the processes that are necessary consequences of carrying out the task (Pressley et al., 1985). Strategies are particularly interesting to understand mechanisms of performance in a task: they describe the particular set of cognitive operations used by a subject, reflecting how cognitive abilities are implemented in a task to create performance. In other words, understanding strategies is a way of understanding the cognitive mechanisms involved in producing a response.

Raven's matrices require subjects to identify the missing piece in a picture among several alternatives, based on logical rules which apply across both rows and columns. This allows two major strategies (for a review, see Laurence & Macedo, 2023): constructive matching, which requires understanding the logical rules to mentally reconstruct the missing piece of the matrix; and response elimination, which involves examining all response alternatives to select one that seems to fit well with the matrix (Snow, 1978, 1980; see also Bethell-Fox et al., 1984). The balance between these two strategies is reflected in gaze patterns in the task: spending a higher proportion of time on the matrix, waiting longer before viewing the responses, and toggling less often between the matrix and the responses all reflect more use of constructive matching and less use of response elimination (Vigneau et al., 2006). Subjects may also use a mix of the two strategies: Jarosz et al. (2019) reported a hybrid isolate-and-eliminate strategy where subjects tried to understand one logical rule, then eliminated response alternatives that did not match this particular rule (see also Li et al., 2022).

Constructive matching is more effective on average, with a greater probability of leading to the correct answer. As a result, strategy used in Raven's matrices and similar tasks is strongly related to performance, and is usually a better predictor of performance in the task than cognitive abilities such as working memory (e.g., Gonthier & Thomassin, 2015; Jarosz et al., 2019). Besides, strategies can shift throughout the task, along with performance (Gonthier & Roulin, 2020), and depend on conditions such as intellectual disability (Vakil et al., 2011; Vakil & Lifshitz-Zehavi, 2012). Strategy use is thus relevant to understanding how subjects can manage to achieve quantitatively high performance in the task, and constitutes a possible avenue of cognitive development (Perret, 2015).

Moreover, strategy use depends on cognitive abilities: using constructive matching requires holding logical rules and perceptual features of an item in working memory, and the use of constructive matching is correlated with working memory capacity (WMC; Gonthier & Roulin, 2020; Gonthier & Thomassin, 2015; Jarosz et al., 2019; Jastrzębski et al., 2018; see also Jarosz & Wiley, 2012). Although the exact nature of the relation between WMC and strategy use is disputed (strategy use may be a mediator of the relation between WMC and reasoning performance: Gonthier & Thomassin, 2015; Jarosz et al., 2019; a moderator: Li et al., 2022; or neither: Jastrzębski et al., 2018), all results so far agree that WMC and strategy use are related in the context of Raven's matrices. Given that WMC improves throughout cognitive development (e.g., Cowan & Alloway, 2009; Simmering & Perone, 2013), the relation between effective strategy use and WMC makes strategy use a viable candidate to explain how cognitive development could result in higher performance in Raven's matrices.

Development of Reasoning Strategies

A substantial corpus of literature has shown that strategy use tends to improve over development. This is true in a variety of areas (e.g., Siegler, 2016); the most studied are probably mathematical cognition (e.g., Lemaire & Siegler, 1995) and memory (e.g., Jarrold, 2017), but there are examples in many other fields from locomotor strategies (Adolph, 1997) to cognitive control (Gonthier et al., 2019) to reading (Paris & Oka, 1986). Strategies matter to the point where cognitive development can largely be viewed as a question of strategy changes (Siegler, 2000). Studies regarding the development of mathematical strategies have led to the creation of the overlapping waves model (Siegler, 2000, 2016; Siegler et al., 1996), which proposes that multiple strategies coexist at all ages, but children gradually progress toward more effective strategies while the frequency of less effective strategy progressively decreases. Overall, strategy use seems to improve across four major dimensions: new strategies are discovered, effective strategies are used more often, children choose the most efficient strategy more adaptively, and execution of strategies improves (Lemaire & Siegler, 1995; Siegler, 2000).

For all the data that have been gathered concerning the development of effective strategy use, there has been surprisingly little interest into the development of strategies in reasoning tasks. An important role has been ascribed to strategies in the development of formal reasoning for children 4–5 aged years and older (e.g., Ricco, 2015), with examples in transitive reasoning (Halford et al., 1995) or causal reasoning (e.g., Chen & Klahr, 1999; Kuhn et al., 1995), but these results cannot be leveraged to understand performance in general intelligence tests, which do not typically use formal reasoning tasks. Likewise, a few authors have studied cognitive tasks (e.g., sudoku-like problems in 8–11-year-olds: Perret et al., 2011; or discrimination learning in 6–10-year-olds: Whitebread, 1996) too specific to be applicable to more usual tasks such as Raven's matrices.

One of the rare examples of developmental studies regarding strategy use in intelligence tests (Rozencwajg & Corroyer, 2001; Rozencwajg et al., 2005) was interested in Kohs' block design task, a classic component of the Wechsler scales requiring subjects to reproduce an abstract design using colored cubes. The authors found three strategies in the block design task: a global strategy of assembling the blocks by trial-and-error, an analytic strategy of reproducing the design in rows or in columns, and a synthetic strategy of reproducing gestalts composed of multiple blocks. Twelve-year-olds (and older adults: Rozencwajg et al., 2005) used the least effective global strategy to a greater extent, whereas 17-year-olds and young adults turned to the more effective analytic and synthetic strategies (Rozencwajg & Corroyer, 2001). This and similar results serve to suggest that there may be a general trend toward more effective strategy use in fluid reasoning tasks throughout cognitive development.

Development of Reasoning Strategies in Matrix Reasoning Tasks

Data regarding the development of effective strategy use in matrix reasoning tasks like Raven's SPM are comparatively very limited, although strategy changes have been listed as a potential contributor to developmental increases of performance in this context (Perret, 2015). Data regarding response times (RTs) in the 6–13 years age range indicate that older children spend comparatively more time on more difficult problems, which could conceivably reflect greater use of constructive matching (Perret & Dauvier, 2018; see also Foorman et al., 1985). However, general RTs are not specifically diagnostic of strategy use, and there are other possible explanations for this pattern, such as developmental improvements of metacognition.

Several developmental studies have been interested in analogy tasks ("A is to B as C is to?"), which are very similar to matrix tasks, although they are less complex and afford other strategies (partly because of their usual focus on semantic relations less prominent in matrix tasks, and partly because of the fact that logical rules apply only on rows, not columns; see e.g., Starr et al., 2018). A series of studies found that 4–6 years old children paid less attention to the logical rule to be deduced (the "A is to B" relationship; Starr et al., 2018), looked at response options earlier in a trial (Thibaut & French, 2016), and were much more affected by the presence and the number of plausible distractors than older children and adults (Guarino et al., 2022; Thibaut et al., 2010a, 2010b). Attention to a plausible distractor was detrimental to accuracy in children (Guarino et al., 2022), and drawing the attention of 5–6-year-olds children to the logical rule to be deduced ("A" and "B") improved their performance (Glady et al., 2017). All these findings could reflect the lack of constructive matching and the comparatively greater use of response elimination in young children.

Early developmental research with matrix-like tasks was in line with these results, finding that children around 6 years old were more sensitive to perceptual factors than older children (Inhelder & Piaget, 1964; Odom et al., 1975; Overton & Brodzinsky, 1972; Smedslund, 1964): for example, children were less successful at solving problems where the dimensions (color, shape, etc.) relevant to the answer were less perceptually salient. At least one study found that children in this age range have a strong tendency to select incorrect responses duplicating an entry already present in the matrix (Siegler & Svetina, 2002). In both cases, this could reflect greater use of response elimination leading to select an answer with perceptual similarity to the matrix, but this conclusion remains speculative and other interpretations are possible—such as difficulty abstracting logical rules form perceptual regularities in young children, or a fallback on categorization tasks trained at elementary school.

Two datasets have provided more direct information regarding use of constructive matching and response elimination in cognitive development. One study used eye-tracking in a matrix task with a small sample of 69 children and found that 7–8-year-olds demonstrated more structured visual exploration of the matrix than 5–6-year-olds, as reflected in fixations of consecutive elements in the same row or column (Chen et al., 2016). This suggests the possibility of greater constructive matching. On the other hand, older children did not spend more time on the items, and they made more toggles between matrix and responses, making the pattern ambiguous; more usual indices of strategy use were not reported in this study. The preprint study of Niebaum and Munakata (2021) found converging results in a group of 81 children, with more structured visual exploration and a lower toggle rate in 9-year-olds than in 6-year-olds; but 6-year-olds were close to chance performance and indices of strategy use were mostly unrelated to performance in the group of 9-year-olds, raising the question of whether strategy use can be adequately measured in these age groups.

All in all, these preliminary data are sufficient to hint at relatively more use of constructive matching for older children: in other words, a change in the relative frequencies of strategies—one of the four aspects of developmental variation in strategy use (Lemaire & Siegler, 1995; Siegler, 2000). However, this requires more systematic examination, as no data set has directly measured constructive matching in a sample with a large age range. This also leaves several major questions entirely unanswered. A major question is whether a developmental increase in the use of constructive matching would be sufficient to explain the massive increase of reasoning performance with age. This requires establishing the similarity between the developmental trajectories of performance and strategy use, and conducting a direct test of strategy use as a mediator of the effect of age on performance.

Other unanswered questions relate to two other aspects of developmental variation in strategy use (Lemaire & Siegler, 1995; Siegler, 2000): strategy execution and strategy adaptivity. Regarding strategy execution, the question is whether the relation between strategy use and performance changes with age. Constructive matching is generally more effective, but response elimination is less costly, and faster to implement for very difficult items, which makes it comparatively efficient when there is a large discrepancy between subject ability and item difficulty (see Gonthier & Roulin, 2020). More generally, children often demonstrate utilization deficiencies, whereby they do not actually benefit from using an effective strategy (Bjorklund et al., 1997; Clerc et al., 2014). These points suggest that the relation between strategy use and performance could be different in young children who may lack the ability to implement constructive matching correctly: for example, using constructive matching may fail to improve the performance of younger children, or the use of response elimination may even be beneficial in younger age groups (Niebaum & Munakata, 2021, tested this possibility but found no evidence in its favor).

Regarding strategy adaptivity, the question is whether there are developmental changes in the ability to adjust strategy use to the difficulty of problems. As noted above, response elimination becomes comparatively efficient for very difficult problems where using constructive matching can take a long time without necessarily leading to the correct answer. In adults, strategy adaptivity is reflected in a progressive decrease in the use of constructive matching as difficulty of the task increases (Gonthier & Roulin, 2020), but it is unknown if children also tailor strategy use to changes of difficulty. Preliminary data suggest that this may be the case, although with a pattern different from prior results in adults: the preprint study of Niebaum and Munakata (2021) reported that toggle rate decreased and structured visual exploration (visual scanning of adjacent entries in the matrix) increased with difficulty, which could reflect the presence of strategy adaptivity in young children. Moreover, the study of Perret and Dauvier (2018) found that older children increased their RTs for difficult problems to a greater extent than younger children (see also Gonthier, 2023b), which could indicate improvements in the ability to adjust the use of constructive matching to changing task demands.

Lastly, a secondary question was whether improvements in effective strategy use could be related to improvements in cognitive abilities (e.g., in 10–12-year-olds, see Imbo & Vandierendonck, 2007). As noted above, the use of constructive matching over response elimination is related to WMC in adults (Gonthier & Roulin, 2020; Gonthier & Thomassin, 2015; Jarosz et al., 2019; Jastrzębski et al., 2018), and WMC improves substantially with age (e.g., Cowan & Alloway, 2009; Simmering & Perone, 2013). This suggests that improvement of effective strategy use in the SPM may be driven by maturation of WMC, with a progressive increase in WMC making it easier for children to mentally combine all rule tokens and perceptual elements to construct the correct answer. Similar results have been found in other areas (for instance, Gonthier et al., 2019, found that the developmental increase in the use of the effective mechanism of proactive control in the AX-CPT, a cognitive control task, was related to increases in WMC), but no data exist regarding the relation between WMC and matrix strategy use in development.

Rationale for the Study: Development of Strategy Use in Raven's Matrices

The overarching goal of this study was to examine the development of effective strategy use (constructive matching vs. response elimination) in a large sample of participants completing Raven's matrices. This included five related questions:

1. What is the developmental trajectory of effective strategy use in Raven's matrices? We expected older participants to use the effective strategy of constructive matching to a greater extent.
2. Does a developmental increase of constructive matching provide a plausible mechanism to explain developmental improvement in reasoning performance? We expected strategy use to partly or fully mediate the effect of age on performance in Raven's matrices.
3. Is the effect of strategy use constant throughout development, or could it be the case that constructive matching fails to benefit younger children? We tested whether the effect of strategy use on performance was moderated by age—in other words, whether it varied from one age group to the next.
4. Do children become progressively more adept at adapting their strategy use to problem difficulty? We expected the effect of item difficulty on strategy use to be moderated by age, possibly because older children increased the use of constructive matching on more difficult problems to a greater extent than younger children.
5. Are developmental changes in strategy use related to WMC? To explore this possibility, we had a subsample of participants[1] complete a second testing session with a WMC measure, and we tested whether WMC mediated the effect of age on strategy use.

Raven's SPM can be used starting around 6 years old, and children younger than 6–7 years old typically fail to process logical relations in a matrix correctly (Chen et al., 2016). Most of the developmental increase occurs between approximately 6 and 14 years old (e.g., J. Raven, 2000), but we wanted our study to also be able to make the connection between developmental data and the young adult samples of students used in most research with Raven's matrices. For these reasons, we measured strategy use in the SPM in a cross-sectional sample recruited in schools, with age groups every 2 years between 6 and 18 years old (first grade, third grade, fifth grade, seventh grade, ninth grade, 11th grade, and first year of university).

There are three major solutions to measure strategy use in Raven's matrices. Self-report questionnaires require subjects to have metacognitive insight into the strategies they use, which is notoriously difficult for young children (Schneider, 2008).[2] The second option is verbal self-reports, such as think-aloud protocols (Jarosz et al., 2019), which need to be rescored by independent raters, making it difficult to collect a large sample size. The third option is eye-tracking, which has very high correlations with performance (predicting about 40% of variance in Raven's matrices in Hayes et al., 2011), but is also costly to use with large samples of children. In this study, we elected to use a behavioral measure derived from eye-tracking (Rivollier et al., 2021; see also Mitchum & Kelley, 2010, for a similar procedure). Raven's matrices were split horizontally; subjects could view either the matrix or the response bank, and could toggle freely between one and the other. This split-screen procedure makes available the same core indices of strategy use obtained in eye-tracking: proportion of time on matrix, latency to first toggle, and toggle rate. All three indices have high correlations with performance (e.g., correlations in the .40–.50 range in Vigneau et al., 2006; see also Laurence et al., 2018).

WMC was measured using the Adaptive Composite ComplEx Span (ACCES), a battery of complex span tasks (Conway et al., 2005; Redick et al., 2012) normed in French and designed specifically for developmental comparisons (Gonthier et al., 2018). The ACCES is an adaptive measure, which means difficulty is automatically adjusted as a function of participant performance; this makes it possible to use the same task for participants from 8 years old to 18 years old without

[1] Our initial intention was to collect WMC in all participants, but the onset of the COVID-19 pandemic made it impossible to return to the schools for the second testing session in some classes. In the end, WMC was available for approximately half the participants in each age group. Because the data were missing strictly at random and because sample size was approximately equal across age groups, we elected to analyze these data as planned.

[2] As a pretest for the current study, we anecdotally tried to use a strategy questionnaire with Raven's matrices in a class of first graders; children expressed difficulty in answering the questions and contrary to data in adults, their responses were uncorrelated with their actual performance in the task.

a floor or ceiling effect in any group. The task is heavily based on reading and is therefore not suitable for 6-year-olds (first grade), which means no WMC data were collected for this age group.

Method

Statistical Power

The effect size of strategy use on performance in Raven's matrices is typically large in adults. Rivollier et al. (2021), using the same method as the current study, found relations between 0.55 and 0.70 between the various strategy measures and performance. A power analysis using G*Power 3.1.9.7 (Faul et al., 2007) showed that the lowest effect size of $r = .55$ required a sample size of $n = 33$ subjects per group to achieve 0.95 power. Given the uncertainty in the size of correlations that could be observed in children, we planned data collection for a conservative sample size of 60 children per age group, or about twice this number (corresponding to three classes for the younger age groups). This would be sufficient to achieve 0.90 power even for the lower effect sizes obtained in adult samples with eye-tracking (Vigneau et al., 2006: lowest correlation $r = .41$) or questionnaires (Gonthier & Thomassin, 2015: lowest correlation $r = .40$).

Participants

Participants were recruited through their schools: two elementary schools, two middle schools, two high schools, and psychology students at the local university took part in the experiment between 2020 and 2022. The schools were selected to represent a balanced mix of socioeconomic levels as much as possible.[3] We invited participation in a number of classes every two levels (elementary school: first grade, third grade, and fifth grade; middle school: seventh grade and ninth grade; high school: 11th grade; university: first year), with the objective of collecting 60–80 participants per age group between 6 and 18 years old. The French school system has widely different class sizes for different grades, which made it difficult to plan for strictly equal sample sizes.

All children in a class were invited to complete the study, and data were collected for all participants for whom consent was obtained, excluding those with a native language other than French, intellectual disability, autism spectrum disorder, visual disability, and dyspraxia (these children were tested but the data were not recorded). Three children were excluded after data collection because of having both total accuracy and average RTs in the SPM in the bottom 2.5% of their age group, suggesting they failed to engage with the task.

The final sample included a total of 474 participants (first grade: $n = 52$, third grade: $n = 67$, fifth grade: $n = 65$, seventh grade: $n = 57$, ninth grade: $n = 78$, 11th grade: $n = 76$, university first year: $n = 79$). Demographic composition of the sample is summarized in Table 1 (participants were asked about their sex, male vs. female, and birthdate; ethnicity is illegal to collect in France and was not recorded). Due to the composition of psychology courses, the group of university first-year students had a very imbalanced sex ratio, but controlling for sex as a covariate did not change any of the results.

As described above, at least half of all subjects per group (50%–75%) completed a second testing session with WMC measurement, with two exceptions. No data were collected for the 6-year-old age group, for whom the WMC task was not suitable; and data were collected for the whole 18-year-old age group, which was downsampled to the size of the next-largest group by random removal to avoid biasing the analyses. A total of 264 subjects with WMC data were thus included in the analyses (see Table 1).

The study was approved by the local Board of Education (September 16, 2019). The experiment was performed in agreement with the heads of participating schools and with the teachers of participating classes. Written informed consent was obtained from the legal guardians of all participating children; all participants additionally provided verbal assent. Individual participant results were not shared with the schools or families.

Materials

Raven's SPM

The version of Raven's matrices used for this study was the SPM (J. Raven et al., 1998), which is suitable for all ability levels from young children to young adults. The SPM includes 60 items divided into five sets of 12 items. Difficulty increases within each set, and from one set to the next (so that item 1 of set B is easier than item 12 of set A but harder than item 1 of set A). As a way to limit testing time, we did not use set A, which only includes items with simple pattern completion: these items are very easy even for young children, providing little information on reasoning performance (e.g., Langener et al., 2022), and they can be solved with a simple visual strategy (see Hunt, 1974 and Gonthier & Roulin, 2020). Participants completed the first item of set A as training, then the 48 items of sets B–E. This was sufficient to obtain very reliable measures of accuracy ($\alpha = .94$) and RTs ($\alpha = .90$).

Strategy Measurement in the SPM

Each matrix in the SPM was presented with the split-screen method described in Rivollier et al. (2021; see their Figure 1): the items were cut in half and participants viewed either the matrix in the top half of the screen, or the bank of response options in the bottom half of the screen. Presentation of an item always started with the matrix. Participants could click on an arrow displayed on the right side of the screen to toggle freely between displaying the matrix and displaying the responses.

We recorded three main measures of strategy use (see Rivollier et al., 2021), directly analogous to the measures used in eye-tracking studies (e.g., Vigneau et al., 2006), and all with high reliability in this data set. The first measure was the proportion of time spent looking at the matrix, versus the responses ($\alpha = .95$). The second measure was the time elapsed before the first toggle to the responses ($\alpha = .91$); contrary to the raw measure of time used in adults but in line with developmental literature, we also computed this measure as a proportion of total time (time to first toggle/total time on the item; $\alpha = .95$) to account for developmental differences in RT (see e.g., Kliegl et al., 1994). The third measure was the total number of toggles between matrix and responses ($\alpha = .91$), which we recoded as a toggle rate (total toggles/total time on the item, in seconds; $\alpha = .93$) to account for differences in RT, in line with adult literature.

[3] Based on national socioeconomic data (https://data.education.gouv.fr/explore/?sort=modified&q=ips), the two elementary schools were at q 12.5 and 80.6, the two middle schools were at percentile 39.4 and 89.5, and the two high schools were at percentile 14.3 and 79.9. All schools were in urban areas. Classes of each level were invited in each of the two elementary, middle and high schools (e.g., it was not the case that all first graders were recruited in only one of the two elementary schools).

Table 1
Demographic Composition of the Sample in Each Age Group

Age group	Grade	N		Female (%)	Male (%)	Age range	Age mean (*SD*)
		SPM	WMC				
6	First	52	—	48	52	5.93–7.14	6.40 (0.30)
8	Third	67	44	54	46	7.07–9.16	8.42 (0.38)
10	Fifth	65	32	55	45	9.29–11.22	10.42 (0.32)
12	Seventh	57	29	51	49	11.36–14.19	12.71 (0.61)
14	Ninth	78	58	44	56	13.01–15.33	14.53 (0.43)
16	11th	76	43	62	38	15.41–18.35	16.76 (0.51)
18	University first year	79	58	90	10	17.32–22.07	19.14 (1.06)

Note. SPM = standard progressive matrices; WMC = working memory capacity.

WMC Measurement With the ACCES

Working memory was measured using the ACCES task (Gonthier et al., 2018; see also Gonthier et al., 2016). The ACCES is a battery of three classic complex spans: the reading span, symmetry span, and operation span (Conway et al., 2005; Redick et al., 2012), presented in this order. All three tasks were adapted from English-speaking adult versions (Unsworth et al., 2005), with simpler concurrent processing demands suitable for children.

The structure of the three tasks is similar and interleaves the presentation of to-be-remembered items with concurrent processing demands (memorize a stimulus, solve a processing task, memorize a stimulus, etc.). In the reading span, participants have to memorize digits while deciding if sentences are correct or not. In the symmetry span, they have to memorize spatial locations presented in a 4 × 4 grid while deciding if pictures are vertically symmetrical or not. In the operation span, they have to memorize consonants while deciding if math operations are correct or not. At the end of a trial, participants are required to recall all to-be-remembered stimuli, in serial order.

Contrary to common adult versions, the ACCES uses an adaptive procedure. Each complex span starts with a moderate set size, and difficulty progressively increases or decreases in subsequent trials depending on participant performance (e.g., a participant recalling five out of five stimuli in a trial is then required to memorize six stimuli). Performance was scored using the edit-distance scoring method, an improved variant of partial-credit scoring with better psychometric properties, especially for low-performing participants and for difficult trials (Gonthier, 2023a; the results did not change when using partial-credit scoring). In each trial, performance was scored depending on the number of changes required to edit the participant's response into the correct sequence (e.g., recalling BADE instead of ABCDE was scored 3 out of 5, because the sequence BADE requires two changes: inverting A and B, and adding a C). Total scores in the three complex span were standardized, then averaged to yield a total WMC score. The ACCES demonstrated a satisfying test–retest reliability ($r = .70$) and a correlation with Raven's matrices very close to expected values ($r = .34$, respectively) in its initial validation on 8–13-years-old children (Gonthier et al., 2018).

Procedure

Children were tested in groups of up to six participants for the 6-, 8- and 10-year-old groups, and up to 10 participants for the older age groups. Data collection took place in a quiet room at the participants' school. An experimenter was present throughout all testing sessions to ensure that participants remained focused on their task. For the 6-year-old and 8-year-old age groups, task instructions were read out loud at the beginning of the task to ensure full understanding. After a participant completed the task of interest, the task script transitioned into a small game that kept the child busy until the end of the testing session, which was discontinued for all participants simultaneously. Each of the two testing sessions took approximately 30–35 min.

Transparency and Openness

This study was not preregistered. All the data are available on the Open Science Framework at https://osf.io/zyemb/ (Gonthier, 2022). The ACCES task used to measure WMC is also available at https://osf.io/bk7pm/ (Gonthier, 2021). Raven's progressive matrices are not made available because of copyright.

Results

Descriptive statistics for all variables as a function of age group are displayed in Table 2. More detailed tables with skewness, kurtosis, and reliability coefficients at the age group level are also available on https://osf.io/zyemb/. Reliability coefficients were above .75 for all measures, and were in the .83–.96 range for all indices of strategy use in all age groups. Skewness and kurtosis were acceptable overall, with a few deviations from normality because of moderately outlying values, but removing the corresponding subjects did not change the results.

Bivariate correlations between all variables are displayed in Table 3. Indices of strategy use all correlated with accuracy on Raven's matrices, although to variable extents: proportion of time on matrix was the best predictor with 53% of explained variance, followed by proportion of latency to first toggle with 22% of explained variance, and toggle rate with 3% of explained variance. Overall, indices of strategy use explained 57% of variance in performance, establishing their predictive validity.

Question 1: Does Effective Strategy Use Increase With Age?

Our data collection allowed age to be treated as categorical (by considering separately the seven age groups: treating age as categorical takes into account a possible nonlinear effect of age) or continuous (because of the age variability within each age group, the older participants in one group were about contiguous to the younger participants in the next: treating age as linear made for simpler analyses). For completeness, we conducted the main analyses of developmental trajectories with both solutions (analyses of variance and regressions).

Table 2
Descriptive Statistics for All Measures as a Function of Age Group

Measure	Age group							Main effect of age group		
	6	8	10	12	14	16	18	F	p	η_p^2
Total accuracy	11.27 (5.77)	17.21 (7.82)	25.08 (7.21)	27.19 (7.64)	33.59 (5.91)	35.29 (4.81)	36.23 (4.98)	93.61	<.001	.53
RT	16.11 (7.85)	17.03 (7.63)	16.72 (6.87)	16.23 (6.73)	18.83 (6.65)	18.11 (4.92)	17.42 (6.02)	1.49	.192	.02
Proportion of time on matrix	0.59 (0.06)	0.62 (0.07)	0.65 (0.06)	0.67 (0.06)	0.68 (0.05)	0.71 (0.04)	0.74 (0.04)	58.32	<.001	.43
Time to first toggle	8.16 (4.20)	9.13 (5.09)	8.41 (3.54)	8.64 (3.39)	10.77 (3.97)	11.09 (3.76)	11.18 (4.06)	7.38	<.001	.09
Proportion of time to first toggle	0.53 (0.07)	0.54 (0.08)	0.54 (0.07)	0.57 (0.11)	0.59 (0.08)	0.63 (0.08)	0.66 (0.08)	26.99	<.001	.26
Total toggles	1.43 (0.44)	1.62 (0.46)	1.93 (0.63)	1.80 (0.76)	1.87 (0.68)	1.72 (0.62)	1.60 (0.50)	5.29	<.001	.06
Toggle rate	0.13 (0.05)	0.14 (0.04)	0.16 (0.05)	0.16 (0.04)	0.14 (0.04)	0.13 (0.04)	0.15 (0.03)	4.45	<.001	.05
WMC	—	−1.41 (0.74)	−0.78 (0.82)	−0.25 (0.69)	0.39 (0.67)	0.53 (0.60)	0.61 (0.59)	70.77	<.001	.56

Note. The table displays means with *SD*s in parentheses. All measures (except WMC) are averages across all items of the SPM. RT and latency to first toggle are in seconds, WMC is the standardized edit distance averaged across the three complex spans. RT = response time; WMC = working memory capacity.

Treating age as a categorical variable, there was a significant effect of age on all measures of performance and strategy use collected here, except for average RT. This is summarized in Table 2. The developmental trajectories of the four main measures—total score on the SPM, proportion of time on matrix, proportion of time to first toggle to the response alternatives, and toggle rate—are displayed in Figure 1.

As usual, accuracy improved with age. The developmental trajectory of performance was reasonably close to linear, although we found the expected plateau starting on about 14-year-old (e.g., J. Raven, 2000). More importantly, there was also a relatively linear increase with age for two of the major measures of strategy use: proportion of time on matrix, and time (or proportion of time) before first toggle. This was reflected in significant linear correlations, as summarized in Table 3. These results indicated that use of the effective strategy of constructive matching increased with age, as predicted.

Total number of toggles and toggle rate were also affected by age, but their developmental trajectory was decidedly nonlinear (for toggle rate, this is represented in Figure 1). There was a progressive increase of toggles between matrix and response alternatives, with a peak around 10-year-old, then a progressive decrease. Despite a slight uptick in the 18-year-old group, average values for the older participants were similar to the 6-year-old group. This resulted is a nonsignificant linear correlation with age (as visible in Table 3). Even when treating age as a categorical variable, the effect of age on total number of toggles and toggle rate was very limited, explaining about 5% of variance. The markedly different developmental trajectory, along with the limited correlation with other strategy indices and prior results using the same design in adults (Rivollier et al., 2021), converged to suggest that toggles may be sensitive to other mechanisms than constructive matching, at least in the context of the specific paradigm used here (but see also Chen et al., 2016). We return to this point in the discussion.

Question 2: Can Effective Strategy Use Explain the Effect of Age on Reasoning?

The developmental trajectories of reasoning performance, and constructive matching as indexed by proportion of time on matrix and proportion of time to first toggle, were very close together, as represented in the top left quadrant of Figure 1. Averaging results at the group level ($n = 7$ age groups), performance was correlated $r = .97$ with proportion of time on matrix and $r = .86$ with proportion of time to first toggle (or $r = .84$ with raw time to first toggle). This made constructive matching a plausible mechanism to explain the developmental improvement of reasoning performance.

The effect of strategy use in relation to age was examined in a multiple regression analysis,[4] with accuracy in the SPM as a dependent variable and age and strategy indices as predictors. This was complemented with a commonality analysis (performed with package yhat, Nimon et al., 2008; for R: R Core Team, 2016), which quantifies the shares of explained variance that are unique or common to multiple predictors (Mood, 1971; for another example in cognitive development, see Gonthier et al., 2019). Due to the presence of multicollinearity between strategy indices, they were considered together for these analyses.

The results are summarized in Figure 2. In a simple regression, age explained 59.3% of variance in SPM accuracy ($r = .77$, $p < .001$; see Table 3). In a multiple regression, age and strategy indices together explained 72.5% of variance in performance ($p < .001$ for all predictors). Out of this total, 15.7% of variance was uniquely explained by age; 13.5% of variance was uniquely explained by the strategy indices; and 43.3% of variance was explained in common by age and strategy use. In other words, variance explained by age decreased from 59.3% to 13.5% when controlling for strategy use: most of the developmental increase of performance was accompanied by concurrent improvements of strategy use.

Another way to examine whether the effect of age could be driven by strategy use is to test whether strategy use mediates the relationship between age and performance. A multiple mediation analysis (using package lavaan: Rosseel, 2012; with bootstrapped confidence intervals and 5,000 resamples: Preacher & Hayes, 2008) indicated that the effect of age on SPM accuracy was significantly mediated by the strategy indices (total indirect effect estimate = 0.22, $z = 6.39$, $p < .001$). The direct effect of age was however still significant when controlling for strategy use, direct effect estimate = 0.55, $z = 15.22$, $p < .001$.

Question 3: Does the Effect of Strategy Use on Performance Change With Age?

We tested whether the effect of strategy use changed as a function of age group using a series of analyses based on the general linear

[4] Given the difficulty of testing a mediation and obtaining a total indirect effect when the predictor is a multicategorical variable (Hayes, 2022), age could only be treated as a continuous variable for this analysis.

Table 3
Bivariate Correlations Between All Measures

Measure	1	2	3	4	5	6	7	8	9
1. Age	—								
2. Total accuracy	**.77**	—							
3. RT	.08	**.35**	—						
4. Proportion of time on matrix	**.65**	**.73**	**.35**	—					
5. Time to first toggle	**.26**	**.48**	**.83**	**.63**	—				
6. Proportion of time to first toggle	**.49**	**.47**	.00	**.78**	**.48**	—			
7. Total toggles	.05	**.17**	**.44**	−.03	.03	**−.58**	—		
8. Toggle rate	.03	**−.18**	**−.56**	**−.32**	**−.69**	**−.41**	**.29**	—	
9. WMC	**.70**	**.74**	.00	**.48**	.15	**.37**	−.06	−.06	—

Note. $N = 474$ for all correlations except those involving WMC, $N = 264$ for correlations involving WMC. Significant correlations (|.16| or above) are in boldface. RT = response time; WMC = working memory capacity.

model, with age as a categorical variable (to avoid imposing a linear shape to the interaction), a measure of strategy use as a continuous variable, and accuracy on the SPM as a dependent variable. The parameter of interest was the interaction between age and strategy use.

The results are displayed in Figure 3. There was no significant interaction between age and strategy use for the major strategy indices: proportion of time on matrix, $F(6, 460) = 0.66$, $p = .680$, $\eta_p^2 = .01$; proportion of time to first toggle, $F(6, 460) = 0.95$, $p = .457$, $\eta_p^2 = .01$; or toggle rate, $F(6, 460) = 2.08$, $p = .054$, $\eta_p^2 = .03$. A complementary Bayesian analysis was performed to quantify evidence in support of the null (using package BayesFactor: Morey & Rouder, 2018); Bayes factors (BFs) for the interaction between age and strategy use were firmly in favor of the null ($BF_{01} = 28920$ for proportion of time on matrix, $BF_{01} = 596$ for proportion of time to first toggle, and $BF_{01} = 171$ for toggle rate). This suggested that the effect of using constructive matching on performance was relatively constant across age groups, ruling out the possibility that response elimination was more beneficial for younger children.

Of secondary interest, toggle rate had a strong negative association with performance in each age group, suggesting that it did in fact function as an index of constructive matching, even though its developmental trajectory appeared sensitive to other factors (see Question 1).

Question 4: Does Strategy Adaptivity Change With Age?

We tested whether strategy adaptivity changes with age in a series of analyses including age and item difficulty as predictors, and strategy measure as dependent variable. To appropriately model changes across the sequence of items, this was done at the item level, using general additive mixed models allowing for nonlinear effects and including a random intercept at the participant level (for details, see Gonthier & Roulin, 2020; see also Perret & Dauvier, 2018). Due to the nonordered difficulty of items in Raven's SPM, item difficulty was not indexed by item ordinal position as in prior work with Raven's advanced progressive matrices (APM; Gonthier & Roulin, 2020), but by the overall proportion of correct answers on an item[5] (1—proportion correct). These analyses were conducted using package mgcv for R (Wood, 2017) with default options. The results are reported as Fisher's F statistics, p values, and effective degrees of freedom (edf, reflecting the degree of nonlinearity, with edf = 1 for a linear trajectory).

The results are displayed in Figure 4. Overall and regardless of age, variation in the three indices of strategy use indicated that participants gradually turned to increased constructive matching as items became more difficult (Figure 4A). This was reflected in significant effects of item difficulty in all of proportion of time on matrix, $F = 350.31$, edf = 8.68, $p < .001$; proportion of latency to first toggle, $F = 80.91$, edf = 8.48, $p < .001$; and toggle rate, $F = 510.67$, edf = 8.44, $p < .001$. Contrary to prior work with the APM in adults (Gonthier & Roulin, 2020), but in line with prior developmental results with the SPM (Perret & Dauvier, 2018), there was little evidence of participants abandoning constructive matching when items became too difficult, apart from a slight downtick of proportion time on matrix and uptick of toggle rate for items above 0.85 difficulty.

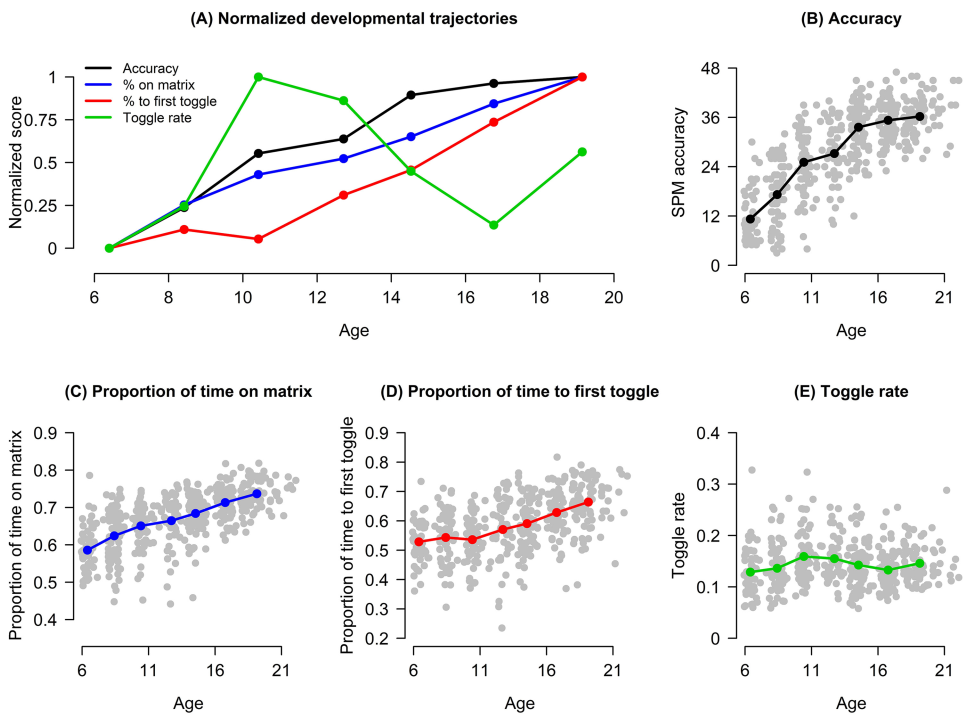
When adding age as a predictor, the interaction between age and item difficulty was significant for all three indices: proportion of time on matrix, $F = 42.156$, edf = 9.61, $p < .001$; proportion of time to first toggle, $F = 82.62$, edf = 12.29, $p < .001$; and toggle rate, $F = 90.60$, edf = 11.69, $p < .001$. As displayed in Figure 3, in all cases older participants had a stronger tendency to turn to constructive matching as difficulty increased, suggesting better strategy adaptivity.

To complement this conclusion and make the connection with prior studies (Perret & Dauvier, 2018; see also Gonthier, 2023b), we performed the same analyses for RTs, with item difficulty and age as predictors. The results are displayed in Figure 5. Consistent with the results for strategy indices, average RTs increased as the difficulty of problems increased (Figure 5A), $F = 492.77$, edf = 8.25, $p < .001$. However, the interaction between age and item difficulty was also significant, $F = 120.07$, edf = 12.31, $p < .001$: younger participants spent approximately the same time on all problems, whereas older participants spent less time on easy items and more time on harder items. This pattern for RTs is broadly consistent with increasing adaptivity of constructive matching use across cognitive development. These results replicated the study of Perret and Dauvier (2018), and additionally confirm why the effect of age on RTs was not significant (see Question 1 and Tables 1 and 2): older participants did not spend longer on matrix problems on average, as they spent less time on easy problems and more time on harder problems.

[5] An alternative solution is to index difficulty as the difficulty parameter of a Rasch model (Perret & Dauvier, 2018). This makes little difference in practice: in this sample, proportion correct and the Rasch difficulty parameter for an item correlated $r = .99$. Difficulty (1—proportion correct) ranged between 0.02 and 0.91, with a median of 0.37 (median absolute deviation = 0.16).

Figure 1

Developmental Trajectories for the Main Measures of Performance and Strategy Use



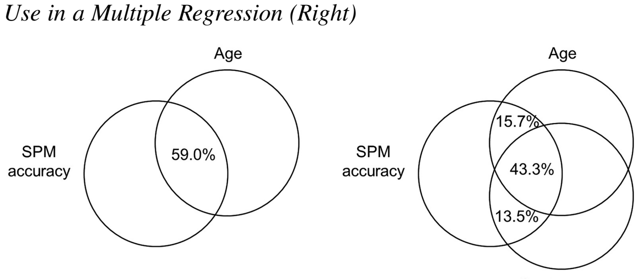
Note. Panel A represents trajectories normalized to be on the same scale (with 0 = *lowest average* and 1 = *highest average score in the sample*). The other panels represent raw trajectories with large dots for age group averages. SPM = standard progressive matrices. See the online article for the color version of this figure.

Question 5: Can WMC Explain the Development of Effective Strategy Use?

The final series of analyses tested the relation between strategy use and WMC in the subsample of participants for whom WMC data were collected ($n = 264$, including all age groups except children in the first grade). There was a significant relation between age and WMC (see Tables 1 and 2), $r = .70$, $p < .001$, as well as between WMC and proportion of time on matrix, $r = .60$, $p < .001$, and between WMC and proportion of time before first toggle, $r = .48$, $p < .001$. This made WMC a possible candidate to explain the age-related increase of constructive matching for these two indices (the mediation for toggle rate was not considered here given the lack of a linear increase with age).

Figure 2

Shares of Variance in SPM Accuracy as Explained by Age in a Simple Regression (Left), and as Explained by Age and Strategy Use in a Multiple Regression (Right)



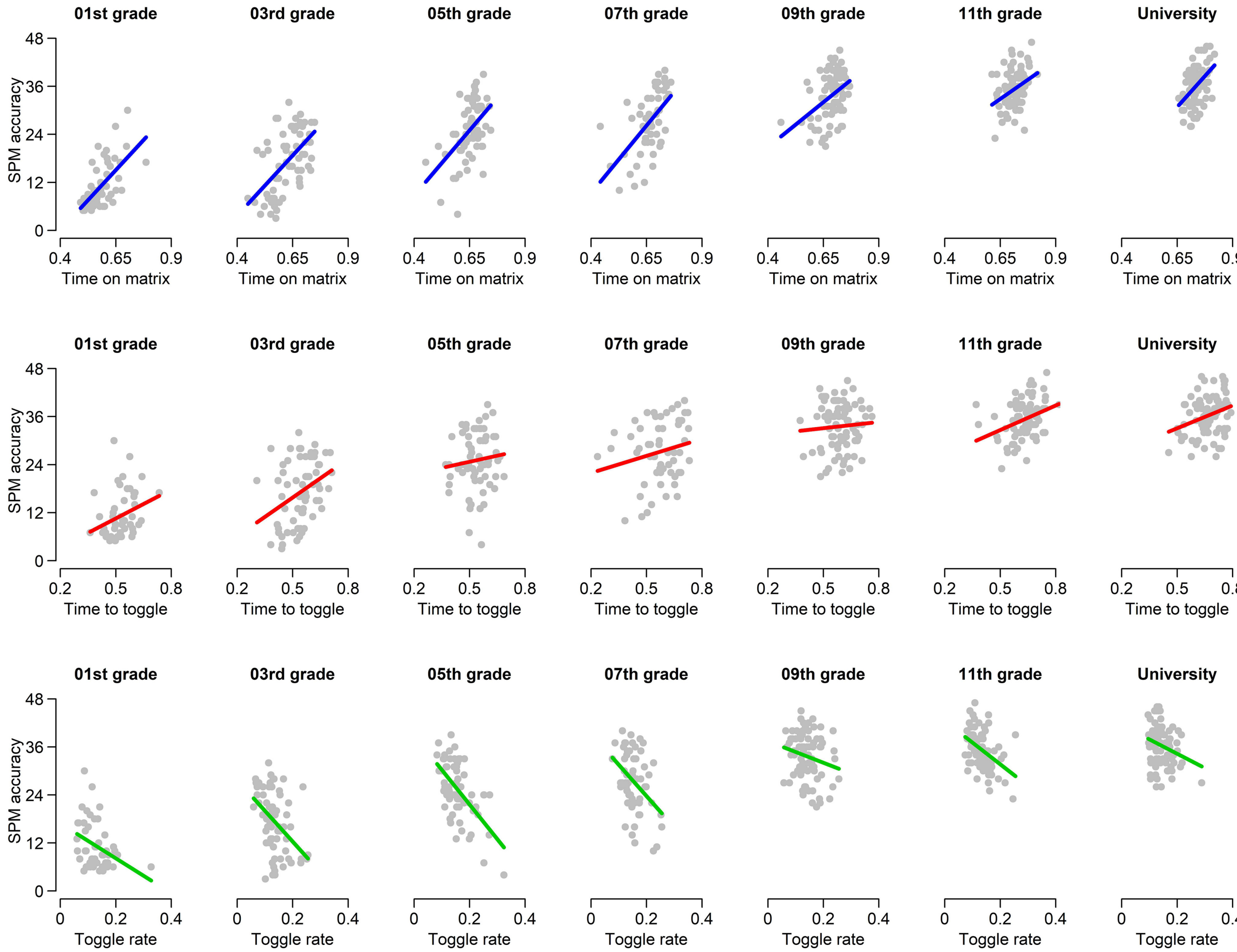
Note. SPM = standard progressive matrices.

The results of multiple regressions with communality analyses are displayed in Figure 6. For proportion of time on matrix, age and WMC together explained 36.6% of variance; out of this total, 13.7% of variance was uniquely explained by age, 0.7% was uniquely explained by WMC, and 22.2% was explained in common by age and WMC ($p < .001$ for age, $p = .067$ for WMC). For proportion of time to first toggle, age and WMC together explained 23.4% of variance; out of this total, 9.4% of variance was uniquely explained by age, 0.3% was uniquely explained by WMC, and 13.7% was explained in common by age and WMC ($p < .001$ for age, $p = .248$ for WMC). In other words, most of the age-related increase in effective strategy use occurred concurrently with age-related improvements of WMC; age retained an effect on strategy use above and beyond WMC; and the effect of WMC on strategy use was negligible when controlling for age. (Of secondary interest, the relation between WMC and strategy use did not interact with age for any index of strategy use, all $ps > .17$.)

Consistent with these results, the mediation by WMC for the effect of age on proportion of time on matrix was only marginally significant with a negligible effect size, indirect effect estimate = .08, $z = 1.72$, $p = .085$. The mediation by WMC for the effect of age on proportion time to first toggle was not significant, indirect estimate = .00, $z = 1.03$, $p = .302$. Overall, these results are compatible with the possibility that changes of strategy occur concurrently with changes of WMC, but not with our hypothesis that WMC could be the key determinant of developmental changes in strategy use.

Figure 3

Correlation Between Indices of Strategy Use and Performance, per Age Group



Note. The first line is for proportion of time on matrix, the second line is for proportion of time to first toggle, and the third line is for toggle rate. SPM = standard progressive matrices. See the online article for the color version of this figure.

Discussion

Our results, representing the first systematic study of developmental variation in effective strategy use on matrix reasoning tasks in a large sample of 474 participants from first grade to the university, led to the following conclusions:

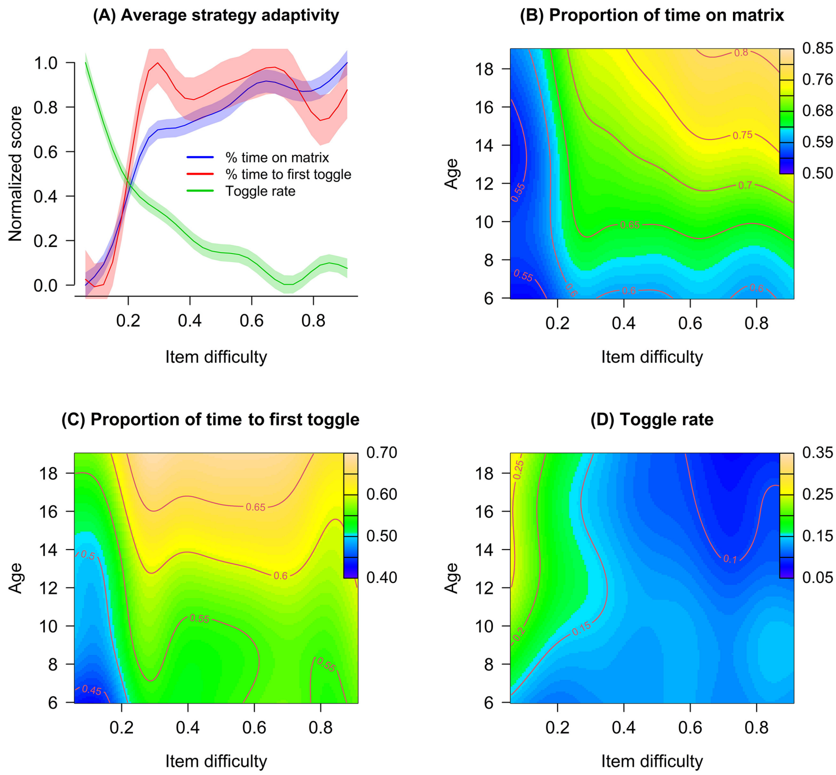
1. Throughout cognitive development, there was a relatively linear increase in use of the effective strategy of constructive matching, as indexed by the proportion of time spent looking at the matrix and time before the first look at the responses, but not as indexed by toggling rate.
2. This increase in effective strategy use accompanied developmental improvements in performance at approximately the same rate, and mediated the effect of age on performance, explaining approximately half of age-related variance.
3. The effect of strategy use on performance did not interact with age, showing that constructive matching was a more effective strategy to a similar extent for participants of all ages.
4. Cognitive development was also accompanied by better strategy adaptivity: younger children spent approximately the same time and used constructive matching to the same extent for all items regardless of difficulty, whereas older participants adjusted their RTs and strategy use to item difficulty.
5. Much of the development of effective strategy use occurred concurrently with increases in WMC, but contrary to our expectations, the development of working memory did not appear to play a predominant role above and beyond age: there was a marginal mediation by WMC, but with a negligible effect size.

Development of Constructive Matching and Reasoning Performance

Overall, the proportion of time spent looking at the matrix and the proportion of time before first toggle showed the expected increase with age, and the expected relations with other measures. This is in line with data from analogy tasks suggesting that young children

Figure 4

Intraindividual Variation for Measures of Strategy Use as a Function of Item Difficulty



Note. Panel A represents changes of strategies throughout the task, averaged across all participants regardless of age, and normalized to be on the same scale. The other panels represent the two-way interaction between item difficulty and age. They are easier to read horizontally: for example in panel B, average proportion of time on the matrix was 0.55 for 6-year-olds on problems below 0.2 difficulty and 0.6 on problems above 0.8 difficulty. See the online article for the color version of this figure.

pay less attention to logical rules (Starr et al., 2018; Thibaut & French, 2016) and are more affected by distracting response options (Guarino et al., 2022; Thibaut et al., 2010a, 2010b), and with data from matrix tasks suggesting that young children tend less to engage in structured visual exploration of the matrix supporting rule inference (Chen et al., 2016; Niebaum & Munakata, 2021). Taken together, and leaving aside the results for toggle rate, these data provide unambiguous evidence that the use of constructive matching increases with age. In sum, older children spend more time on the matrix trying to reconstruct the correct answer.

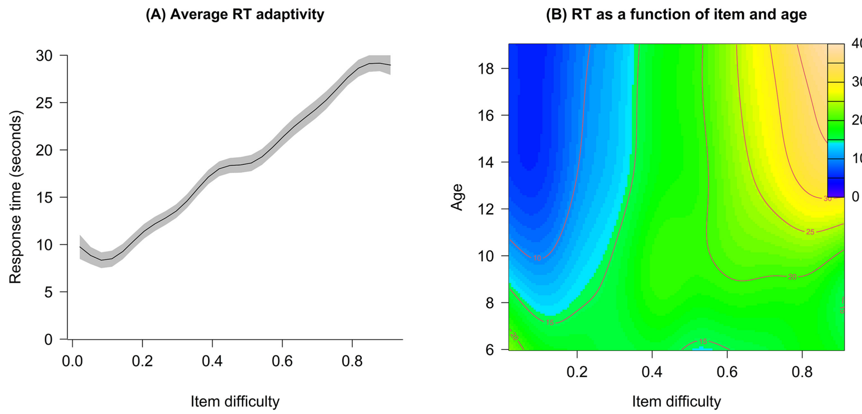
Given that the developmental change of strategy occurred together with a dramatic increase in performance, and given that effective strategy use was related to performance at all ages, the better performance of older children may be partly driven by their more effective strategy use. We found that strategy use partly mediated the effect of age, compatible with this possibility. Importantly, this does not exclude the possibility that the change in strategy use is itself caused by increasing ability: older children might use more constructive matching *because* they are better able to understand logical rules in the matrix, as suggested by the finding of less constructive matching in participants with intellectual disability (Vakil et al., 2011; Vakil & Lifshitz-Zehavi, 2012).

This comes down to asking whether the increase of constructive matching should be viewed as the primary cause of increasingly better reasoning over age, or as the indirect consequence of increasing ability for reasoning. Both could also be true, with ability and strategy use forming a causal loop (see Gonthier & Roulin, 2020). This question is recurring in the study of cognitive strategies. The best way to answer it experimentally would probably be to induce strategy use: if younger children perform lower because they make less use of the effective strategy of constructive matching, then inducing them to use more constructive matching should increase their performance to the level of older children (see Gonthier & Thomassin, 2015). One study with an analogy task provided preliminary evidence in this direction, where 5–6-year-olds performed higher when encouraged to focus their attention on logical rules prior to considering response options (Glady et al., 2017). Another option for supporting constructive matching in children would be to ask them to draw the missing part of the matrix, without providing response options (see Becker et al., 2016; Duncan et al., 2017; Koch et al., 2022)—although this would be very time-consuming in young children.

The issue of causality in the development of effective strategy use indirectly questions the nature of fluid intelligence and its measurement. It has been argued that the existence of strategic variability is a problem for the measurement of intelligence because it implies the test does not measure a single source of variance (Hunt, 1974). This point is debatable: if children turn to a more effective strategy because their ability

Figure 5

Intraindividual Variation for RT (in Seconds) as a Function of Item Difficulty



Note. RT = response time. See the online article for the color version of this figure.

increases, then qualitative changes could partly reflect useful variance in reasoning; besides, being able to select the more effective strategy can also be taken as a meaningful aspect of reasoning. Regardless, the present results unambiguously confirm that the developmental increase in reasoning performance cannot be simply taken to mean that children get "more intelligent" as they grow older, and that cognitive development has to be viewed as a combination of qualitative and quantitative changes that ultimately contribute to increasing performance, not a monolithic increase along the continuum of a unitary ability.

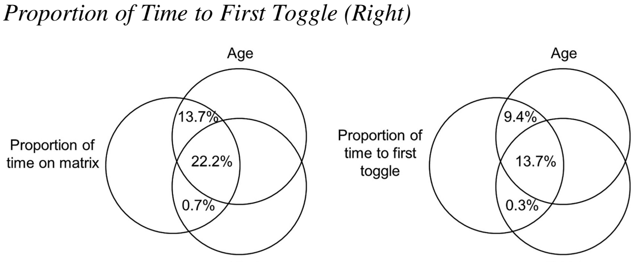
Strategy Adaptivity in the Face of Difficulty

A secondary axis of developmental improvement is strategy adaptivity: older children appeared to be more adept at tailoring the use of constructive matching to problem difficulty, as reflected in increased constructive matching for more difficult problems. These results are especially compatible with those of Perret and Dauvier (2018); see also Gonthier, 2023, showing that older children modulate their RT on difficult problems to a much larger extent; we replicated this finding with our own analysis of RTs (see Figure 5). The conclusion that children become progressively better at selecting a more effective strategy depending on the problem is well in line with the developmental literature (e.g., Lemaire & Reder, 1999).

At first glance, the finding that participants increased constructive matching in the face of difficulty seems contrary to data previously

Figure 6

Shares of Variance Explained by Age and WMC in a Multiple Regression, for Proportion of Time on Matrix (Left) and for Proportion of Time to First Toggle (Right)



Note. WMC = working memory capacity.

reported in adults, which showed less constructive matching for very difficult items, as moderated by ability: all adult participants except those with high WMC and high need for cognition tended to turn to response elimination in the face of increasing difficulty (Gonthier & Roulin, 2020). There was little indication of the same pattern here, apart descriptively from a small downtick of constructive matching on the hardest problems. The 18-year-old group in the current study was similar in composition to the sample of our prior study (Gonthier & Roulin, 2020), so age is not the source of the discrepancy. Instead, it is likely that the difference comes from use of the standard version of Raven's progressive matrices in the current study, versus the advanced version in the prior study.

We previously proposed that strategy changes within the task depend on the perceived benefit and perceived cost of implementing a strategy (Gonthier & Roulin, 2020): participants turn to the strategy that seems most useful and least costly to use. In Raven's advanced matrices, difficult items are very hard for a typical sample of young adults (average performance for the hardest items is below chance level), and come with a high perceived cost to implement constructive matching (there are more elements to combine the reconstruct the correct answer), and a low perceived benefit (low likelihood of success). This presumably leads to disengagement from constructive matching (see also Law et al., 2022). By contrast, in Raven's standard matrices, difficult items are comparatively easier, and require the combination of less elements to reconstruct the correct answer. This gives a better cost–benefit ratio to increasing constructive matching in the face of increasing difficulty.

This difference could be strengthened by the different structures of the two tasks. Raven's advanced matrices monotonically increase in difficulty, which makes it increasingly unlikely that items can be solved through constructive matching. On the other hand, the standard matrices used in the present study alternate easy and difficult items, possibly preserving participant engagement. This is compatible with the finding that use of constructive matching on difficult problems depends on aspects of motivation (need for cognition: Gonthier & Roulin, 2020).

In sum, the present results show that a different pattern of strategy adaptivity should be expected for different versions of the task, depending on problem difficulty (and possibly task structure). Based on the previously proposed pattern of strategy adaptivity depending on perceived costs and benefits, we argue that a progressive increase in constructive matching for difficult items should be expected when difficult items require more engagement but still seem relatively solvable (as in the SPM), but a progressive decrease should be expected when the task becomes disproportionately difficult (as in the APM). Of course, this pattern should be moderated by individual ability and motivation, as previously shown in adults (Gonthier & Roulin, 2020).

Determinants of the Development of Effective Strategy Use

Many processes and abilities develop in the wide age range of 6 to 18-year-old considered here, and could contribute to the developmental increase of constructive matching. The current study investigated working memory as the most promising candidate based on prior literature, with mixed results. On one hand, much of the age-related variance in constructive matching was shared with WMC, indicating that constructive matching and WMC developed at similar rates. This is compatible with the possibility that the increase of constructive matching is partly supported by increasing WMC, allowing for easier integration of logical rules to construct the correct answer. On the other hand, there was only marginal mediation of the relation between age and constructive matching by WMC, with a very small effect size (about 1% of explained variance), and there was little effect of WMC after accounting for age; in other words, WMC only had a limited relation to constructive matching within each age group. This is incompatible with the hypothesis that the age-related increase in constructive matching is primarily *caused* by increasing WMC.

The fact that WMC develops in tandem with constructive matching, but does not retain a major relation with constructive matching when accounting for age, is at odds with prior studies finding a substantial relation between WMC and constructive matching in adults (Gonthier & Roulin, 2020; Gonthier & Thomassin, 2015; Jarosz et al., 2019; Jarosz & Wiley, 2012; Jastrzębski et al., 2018). As was the case for strategy adaptivity, this different pattern may be caused by the use of a different task. Raven's SPM are less complex and tend to include fewer logical rules than Raven's APM: it is possible that the two versions rely on different cognitive abilities to different extents, and that constructive matching in the SPM depends substantially less on working memory.

Our results do not support the hypothesis of a primary causal role of WMC (at least in this age range: increases in WMC could still account for the progressive emergence of the ability to solve items beyond simple pattern completion in children younger than 8-year-old). However, the data shed little light on the determinant or determinants that elicited the developmental increase of constructive matching in our sample. Metacognition could be a possible contributor (see Perret & Dauvier, 2018), although not one that we explored in the present study: older children are typically better at judging their own performance and judging their success rate with various strategies, and they may engage more constructive matching because they have a more accurate notion that it is necessary to obtain high performance (but see Cary & Reder, 2002). This hypothesis would be relatively straightforward to test by collecting self-estimates of item difficulty, or by manipulating the presence of feedback in the task.

In the same way that various abilities may contribute differently to different versions of the task depending on their requirements, it may also be the case that various abilities contribute differently at different ages (e.g., see Demetriou et al., 2014). In particular, the use of constructive matching may be more constrained by the basic ability to perform the task in the youngest children: indeed, children around 5–6 years old tend to perform around chance in complex matrix tasks (Chen et al., 2016; Niebaum & Munakata, 2021), presumably because of difficulty mentally representing relations between elements of the matrix (see Chen et al., 2016). Conversely, use of constructive matching for older children may depend more on ability to infer the rules, metacognition, or yet other abilities.

Measures of Strategy Use in a Developmental Context

The results were unexpected for number of toggles and toggle rate, which showed a complex picture with two phases of developmental progression: toggles increased between 6 and 10 years old, then decreased until late adolescence. Based on the literature, this should be incompatible with the changes observed for proportion of time on matrix and proportion of time before first toggle, which both counterintuitively increased along with toggle rate until 10 years old. It is all the more surprising that toggle rate was overall negatively correlated with the other strategy indices and with performance, as expected for an index of response elimination. We believe this pattern to be because of the fact that toggles are not solely an index of response elimination: they also represent lookbacks to the matrix for verification, both to double-check one's answer and to refresh the memory trace of matrix components. In this light, we believe the diverging trajectory for toggle rate to be because of two processes operating concurrently with developmental changes of constructive matching.

First, the increase in toggles between 6 and 10 years old may be driven by increasing effort on the part of children to double-check their answer. This would be compatible with an increase of constructive matching as reflected in other indices, and this would be in line with the results of Chen et al. (2016), who found that number of toggles increased with age, and that more toggles were associated with better performance in young children. Of course, other reasons are possible, such as children progressively realizing that toggles are beneficial to decrease the need to hold all components of the item in working memory. Second, the decrease in toggles after 10 years old may be driven in part by increase in WMC, facilitating maintenance of elements of the problem in working memory and limiting the need to look back at the matrix when examining responses. Older children may be comparably more adept at maintaining relevant information from the matrix in working memory (or perhaps at selecting relevant information in the first place), decreasing the need to return to check the matrix.

It is unclear, based on this and prior data, to what extent this unexpected trajectory for toggles was caused by the particular paradigm we used to measure strategy use, and to what extent it is an actual feature of cognitive development. Comparatively lower effort in young children to look back at the matrix to confirm a solution could be reinforced by our paradigm: with only half of the item displayed at one time, participants had to make a purposeful effort to move the mouse and click to toggle between the matrix and responses. As in our prior study using the same paradigm with adults (Rivollier et al., 2021), this led to considerably lower number of toggles and toggle rate on average than in comparable studies using eye-tracking (e.g., Chen et al., 2016; Niebaum & Munakata, 2021). Likewise, having only half the item displayed at a given time may have limited quick lookbacks and placed more constraints on working memory to keep elements of the matrix in mind while looking at the responses, and vice versa. On the other hand, it is entirely possible that this unexpected pattern with an early increase in toggles would have been found regardless of the paradigm, given that similar results were reported by Chen et al. (2016) using eye-tracking.

It seems unworthwhile to dismiss toggle rate as an index of constructive matching based on these results: toggle rate had a negative correlation with performance in all age groups, and at least one study found it to be the best predictor of performance compared to other indices of strategy use (Laurence et al., 2018). However, two recommendations can be given. First, indices based on time on matrix and indices based on toggles are often considered interchangeably, but the current results make it clear that they can behave in very different ways (see also Rivollier et al., 2021), may involve different cognitive processes at different ages, and should be treated separately. We argue that relative time on matrix is a more direct reflexion of a subject's attempt to use constructive matching to understand rules and reconstruct the correct answer, and that toggles should be considered as a more composite index incorporating verifications and memory refreshes. Second, the timing at which toggles occur within a trial could help disentangle their meaning: for instance, a toggle occurring quickly after item presentation might reflect response elimination, whereas a series of close toggles immediately before a response might reflect checking one's answer. Recording information on the timing of toggles is not usual and was not done in the present study, but it would be straightforward both with eye-tracking and with our behavioral paradigm, and should be considered in the future (see e.g., Thibaut & French, 2016).

On a secondary note, proportion of time on the matrix before the first toggle is usually computed as a raw time (e.g., Vigneau et al., 2006), whereas we chose to compute it as a proportion, in order to correct for developmental differences in speed. Using raw time or proportion of time to first toggle did not appear to make a major difference in most analyses (save perhaps for strategy adaptivity, where the pattern for raw time to first toggle, not detailed in the Results section, was very similar to that displayed in Figure 5 for overall RT). However, the results showed that raw time before the first toggle was very highly correlated with RT (see Table 3); it also had a high negative correlation with toggle rate, probably inflated because of toggle rate depending on the inverse of RT. This was not the case for proportion of time before first toggle, which also had better relations with age, proportion of time on matrix, total number of toggles, and WMC, supporting the usefulness of this measure for future studies.

A final point of discussion is the possibility that strategies other than constructive matching and response elimination were used. Guessing is a rarely discussed strategy (Gonthier, 2023b), one that did not seem prevalent in the current study given the long RTs observed on average for difficult items. A more problematic point is that the easiest items of the standard matrices require simple pattern completion, and might be solved through a purely visual process of gestalt completion, rather than a dedicated series of constructive matching operations (Hunt, 1974; see also DeShon et al., 1995). We attempted to limit the use of this strategy by removing the first series of 12 items (which are all limited to simple pattern completion), but a few items in the rest of the task can also be solved this way. This could potentially explain part of the increase in constructive matching observed with increasing item difficulty, contrary to the pattern found with the APM (Gonthier & Roulin, 2020). This simple strategy of visual gestalt completion only applies to very easy matrices and has never been systematically studied in the literature. Unfortunately, it would be difficult to estimate the use of this strategy based on eye-tracking or the paradigm used here (it would presumably be associated with fast RTs and a single toggle from matrix to response, but not necessarily with a high proportion of time on the matrix).

Other Directions for Future Studies

A first major extension of our study would be to use a different method to conceptually replicate the results. As discussed above, questionnaires are an option, but one that is ill-suited to young children; verbal self-reports (e.g., think-aloud instructions: e.g., Jarosz et al., 2019) may require less metacognitive insight from children and could conceivably work. Eye-tracking would also be a possible choice. This effort has been started by Niebaum and Munakata (2021) in their preprint study. Eye-tracking is conceptually similar to the split-screen method used here (Rivollier et al., 2021), and although it has the downside of making data collection more difficult for a large sample, it has two major advantages. The first is that eye-tracking allows for more fine-grained analysis of gaze patterns, including detailed analysis of which distractors among the response options are more salient for a participant (Jarosz & Wiley, 2012; see also Guarino et al., 2022), and how series of fixations on the matrix are structured (see especially the matrix time distribution index in Vigneau et al., 2006, and the encoding relations index in Chen et al., 2016). This allows for more systematic analysis of how effectively participants conduct constructive matching and response elimination.

Another advantage of replicating our results with eye-tracking would be that as discussed above, our split-screen method makes toggles more costly and therefore less frequent, and presumably increases working memory load when having to select response options because of greater difficulty looking back at the matrix (Rivollier et al., 2021). It is also possible that having participants start on the matrix, and requiring them to make an effort to toggle to view response options, indirectly induces constructive matching (Gonthier & Thomassin, 2015; Mitchum & Kelley, 2010). This could potentially bias the estimated developmental trajectory of strategy use by overestimating the frequency of constructive matching to some extent.

A second major direction for future studies would be to investigate the rest of the lifespan trajectory of strategy use in matrix reasoning. Fluid reasoning performance substantially increases during childhood, but it also decreases in older age: in the illustrative data set of J. Raven (2000, Figure 5), median SPM performance decreased from 44 out of 60 correct answers at 20 years old to just 24 out of 60 correct answers at 65 years old. It seems likely that these later changes are associated with a corresponding decline in constructive matching use, as is the case for Kohs' block design task (Rozencwajg et al., 2005). Studying strategy use in older adults would help elucidate the rest of the lifespan trajectory for qualitative underpinnings of behavior in the test most frequently used to measure intelligence.

Constraints on Generality

We are confident that the child samples were large enough and representative enough in terms of gender, age, and socioeconomic levels, that the results should generalize to children in general in the 6–18 age range. Given the visuo-spatial nature of the task, there is no reason to expect that the results should be different for children from other high-income countries speaking other languages. The 18-year-old age group was composed of university students in psychology, which slightly oversamples the upper end of the ability distribution in France; performance may differ slightly for nonstudents. The results are not expected to hold for younger children and samples with disabilities, which tend to display different strategies or be unable to perform tasks of this difficulty.

The two main strategies used in Raven's matrices apply to all tasks where the solution can be inferred from the problem or chosen from multiple response options (such as the paper folding test: Bethell-Fox et al., 1984; see also Snow, 1978). Our conclusions regarding the major role of strategy use in cognitive development are expected to hold for all other reasoning tasks with this structure (and given the similar results found with the block design task by Rozencwajg and Corroyer, 2001, possibly for reasoning tasks in general).

References

Adolph, K. E. (1997). Learning in the development of infant locomotion. *Monographs of the Society for Research in Child Development*, *62*(3), 1–140. https://doi.org/10.2307/1166199

Becker, N., Schmitz, F., Falk, A. M., Feldbrügge, J., Recktenwalf, D. R., Wilhelm, O., Preckel, F., & Spinath, F. M. (2016). Preventing response elimination strategies improves the convergent validity of figural matrices. *Journal of Intelligence*, *4*(1), Article 2. https://doi.org/10.3390/jintelligence4010002

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

Binet, A., & Simon, T. (1904). Application des méthodes nouvelles au diagnostic du niveau intellectuel chez des enfants normaux et anormaux d'hospice et d'école primaire. *L'Année Psychologique*, *11*(1), 245–336. https://doi.org/10.3406/psy.1904.3676

Bjorklund, D. F., Miller, P. H., Coyle, T. R., & Slawinski, J. L. (1997). Instructing children to use memory strategies: Evidence of utilization deficiencies in memory training studies. *Developmental Review*, *17*(4), 411–441. https://doi.org/10.1006/drev.1997.0440

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*(3), 404–431. https://doi.org/10.1037/0033-295X.97.3.404

Cary, M., & Reder, L. M. (2002). Metacognition in strategy selection. In P. Chambres, M. Izaute, & P.-J. Marescaux (Eds.), *Metacognition: Process, function and use* (pp. 63–77). Kluwer Academic Publishers. https://doi.org/10.1007/978–1-4615-1099-4_5

Chen, Z., Honomichl, R., Kennedy, D., & Tan, E. (2016). Aiming to complete the matrix: Eye-movement analysis of processing strategies in children's relational thinking. *Developmental Psychology*, *52*(6), 867–878. https://doi.org/10.1037/dev0000113

Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the Control of Variables Strategy. *Child Development*, *70*(5), 1098–1120. https://doi.org/10.1111/1467-8624.00081

Clerc, J., Miller, P. H., & Cosnefroy, L. (2014). Young children's transfer of strategies: Utilization deficiencies, executive function, and metacognition. *Developmental Review*, *34*(4), 378–393. https://doi.org/10.1016/j.dr.2014.10.002

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*(5), 769–786. https://doi.org/10.3758/BF03196772

Cowan, N., & Alloway, T. (2009). Development of working memory in childhood. In M. L. Courage & N. Cowan (Eds.), *The development of memory in infancy and childhood* (2nd ed., pp. 303–342). Psychology Press.

de Ribaupierre, A., & Lecerf, T. (2006). Relationships between working memory and intelligence from a developmental perspective: Convergent evidence from a neo-Piagetian and a psychometric approach. *European Journal of Cognitive Psychology*, *18*(1), 109–137. https://doi.org/10.1080/09541440500216127

Demetriou, A., Spanoudis, G., Shayer, M., van der Ven, S., Brydges, C. R., Kroesbergen, E., Podjarny, G., & Swanson, H. L. (2014). Relations between speed, working memory, and intelligence from preschool to adulthood: Structural equation modeling of 14 studies. *Intelligence*, *46*, 107–121. https://doi.org/10.1016/j.intell.2014.05.013

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. https://doi.org/10.1016/0160-2896(95)90023-3

Duncan, J., Chylinski, D., Mitchell, D. J., & Bhandari, A. (2017). Complexity and compositionality in fluid intelligence. *Proceedings of the National Academy of Sciences*, *114*(20), 5295–5299. https://doi.org/10.1073/pnas.1621147114

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). GPower 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Foorman, B. R., Sadowski, B. R., & Basen, J. A. (1985). Children's solutions for figural matrices: Developmental differences in strategies and effects of matrix characteristics. *Journal of Experimental Child Psychology*, *39*(1), 107–130. https://doi.org/10.1016/0022-0965(85)90032-3

Glady, Y., French, R. M., & Thibaut, J.-P. (2017). Children's failure in analogical reasoning tasks: A problem of focus of attention and information integration? *Frontiers in Psychology*, *8*, Article 707. https://doi.org/10.3389/fpsyg.2017.00707

Gonthier, C. (2021, April 9). *Materials for the CCS and the ACCES* [Open Science Framework data repository]. https://osf.io/bk7pm

Gonthier, C. (2022, November 29). *Development of reasoning performance in Raven's matrices is grounded in the development of effective strategy use* [Open Science Framework data repository]. https://osf.io/zyemb

Gonthier, C. (2023a). An easy way to improve scoring of memory span tasks: The edit distance, beyond "correct recall in the correct serial position. *Behavior Research Methods*, *55*(4), 2021–2036. https://doi.org/10.3758/s13428-022-01908-2

Gonthier, C. (2023b). Should intelligence tests be speeded or unspeeded? A brief review of the effects of time pressure on response processes and an experimental study with Raven's matrices. *Journal of Intelligence*, *11*(6), Article 120. https://doi.org/10.3390/jintelligence11060120

Gonthier, C., Aubry, A., & Bourdin, B. (2018). Measuring working memory capacity in children using adaptive tasks: Example validation of an adaptive complex span. *Behavior Research Methods*, *50*(3), 910–921. https://doi.org/10.3758/s13428-017-0916-4

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 Raven's matrices. *Journal of Experimental Psychology: General*, *144*(5), 916–924. https://doi.org/10.1037/xge0000101

Gonthier, C., Thomassin, N., & Roulin, J.-L. (2016). The Composite Complex Span: French validation of a short working memory task. *Behavior Research Methods*, *48*(1), 233–242. https://doi.org/10.3758/s13428-015-0566-3

Gonthier, C., Zira, M., Colé, P., & Blaye, A. (2019). Evidencing the developmental shift from reactive to proactive control in early childhood and its relationship to working memory. *Journal of Experimental Child Psychology*, *177*, 1–16. https://doi.org/10.1016/j.jecp.2018.07.001

Guarino, K. F., Wakefield, E. M., Morrison, R. G., & Richland, L. E. (2022). Why do children struggle on analogical reasoning tasks? Considering the role of problem format by measuring visual attention. *Acta Psychologica*, *224*, Article 103505. https://doi.org/10.1016/j.actpsy.2022.103505

Halford, G. S., Smith, S. B., Dickson, J. C., Maybery, M. T., Kelly, M. E., Bain, J. D., & Stewart, J. E. M. (1995). Modeling the development of reasoning strategies: The roles of analogy, knowledge, and capacity. In T. J. Simon & G. S. Halford (Eds.), *Developing cognitive competence: New approaches to process modeling* (pp. 77–156). Lawrence Erlbaum Associates.

Hayes, A. F. (2022). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (3rd ed.). The Guilford Press.

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), Article 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

Huepe, D., Roca, M., Salas, N., Canales-Johnson, A., Rivera-Rei, A. A., Zamorano, L., Concepción, A., Manes, F., & Ibañez, A. (2011). Fluid intelligence and psychosocial outcome: From logical problem solving to social adaptation. *PLoS ONE*, *6*(9), Article e24858. https://doi.org/10.1371/journal.pone.0024858

Hunt, E. (1974). Quote the Raven? Nevermore. In L. W. Gregg (Ed.), *Knowledge and cognition* (pp. 129–158). Lawrence Erlbaum.

Imbo, I., & Vandierendonck, A. (2007). The development of strategy use in elementary school children: Working memory and individual differences. *Journal of Experimental Child Psychology*, *96*(4), 284–309. https://doi.org/10.1016/j.jecp.2006.09.001

Inhelder, B., & Piaget, J. (1964). *The early growth of logic in the child: Classification and seriation*. Routledge.

Jarrold, C. (2017). Working out how working memory works: Evidence from typical and atypical development. *Quarterly Journal of Experimental Psychology*, *70*(9), 1747–1767. https://doi.org/10.1080/17470218.2016.1213869

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

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

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

Kail, R. V. (2007). Longitudinal evidence that increases in processing speed and working memory enhance children's reasoning. *Psychological Science*, *18*(4), 312–313. https://doi.org/10.1111/j.1467-9280.2007.01895.x

Kliegl, R., Mayr, U., & Krampe, R. T. (1994). Time-accuracy functions for determining process and person differences: An application to cognitive aging. *Cognitive Psychology*, *26*(2), 134–164. https://doi.org/10.1006/cogp.1994.1005

Koch, M., Spinath, F. M., Greiff, S., & Becker, N. (2022). Development and validation of the Open Matrices Item Bank. *Journal of Intelligence*, *10*(3), Article 41. https://doi.org/10.3390/jintelligence10030041

Kuhn, D., Garcia-Mila, M., Zohar, A., & Andersen, C. (1995). Strategies of knowledge acquisition. *Monographs of the Society for Research in Child Development*, *60*(4), v–128. https://doi.org/10.2307/1166059

Laidra, K., Pullmann, H., & Allik, J. (2007). Personality and intelligence as predictors of academic achievement: A cross-sectional study from elementary to secondary school. *Personality and Individual Differences*, *42*(3), 441–451. https://doi.org/10.1016/j.paid.2006.08.001

Langener, A. M., Kramer, A., den Bos, W., & Huizenga, H. M. (2022). A shortened version of Raven's standard progressive matrices for children and adolescents. *British Journal of Developmental Psychology*, *40*(1), 35–45. https://doi.org/10.1111/bjdp.12381

Laurence, P. G., & Macedo, E. C. (2023). Cognitive strategies in matrix-reasoning tasks: State of the art. *Psychonomic Bulletin & Review*, *30*(1), 147–159. https://doi.org/10.3758/s13423-022-02160-7

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*, Article 380. https://doi.org/10.3389/fpsyg.2018.00380

Law, M. K. H., Stankov, L., & Kleitman, S. (2022). I choose to opt-out of answering: Individual differences in giving up behaviour on cognitive tests. *Journal of Intelligence*, *10*(4), Article 86. https://doi.org/10.3390/jintelligence10040086

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. https://doi.org/10.3758/BF03211420

Lemaire, P., & Siegler, R. S. (1995). Four aspects of strategic change: Contributions to children's learning of multiplication. *Journal of Experimental Psychology: General*, *124*(1), 83–97. https://doi.org/10.1037/0096-3445.124.1.83

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. https://doi.org/10.1016/j.intell.2022.101627

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

Mood, A. M. (1971). Partitioning variance in multiple regression analyses as a tool for developing learning models. *American Educational Research Journal*, *8*(2), 191–202. https://doi.org/10.3102/00028312008002191

Morey, R. D., & Rouder, J. N. (2018). *Bayesfactor: Computation of Bayes factors for common designs*. https://cran.r-project.org/package=BayesFactor

Niebaum, J., & Munakata, Y. (2021). *The development of relational reasoning: An eyetracking analysis of strategy use and adaptation in children and adults performing matrix completion* [Unpublished preprint]. https://psyarxiv.com/wfz8u/

Nimon, K., Lewis, M., Kane, R., & Haynes, R. M. (2008). An R package to compute commonality coefficients in the multiple regression case: An introduction to the package and a practical example. *Behavior Research Methods*, *40*(2), 457–466. https://doi.org/10.3758/BRM.40.2.457

Odom, R. D., Astor, E. C., & Cunningham, J. G. (1975). Effects of perceptual salience on the matrix task performance of four- and six-year-old children. *Child Development*, *46*(3), 758–762. https://doi.org/10.2307/1128575

Overton, W. F., & Brodzinsky, D. (1972). Perceptual and logical factors in the development of multiplicative classification. *Developmental Psychology*, *6*(1), 104–109. https://doi.org/10.1037/h0032204

Paris, S. G., & Oka, E. R. (1986). Children's reading strategies, metacognition, and motivation. *Developmental Review*, *6*(1), 25–56. https://doi.org/10.1016/0273-2297(86)90002-X

Perret, P. (2015). Children's inductive reasoning: Developmental and educational perspectives. *Journal of Cognitive Education and Psychology*, *14*(3), 389–408. https://doi.org/10.1891/1945-8959.14.3.389

Perret, P., Bailleux, C., & Dauvier, B. (2011). The influence of relational complexity and strategy selection on children's reasoning in the Latin Square Task. *Cognitive Development*, *26*(2), 127–141. https://doi.org/10.1016/j.cogdev.2010.12.003

Perret, P., & Dauvier, B. (2018). Children's allocation of study time during the solution of Raven's progressive matrices. *Journal of Intelligence*, *6*(1), Article 9. https://doi.org/10.3390/jintelligence6010009

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*(3), 879–891. https://doi.org/10.3758/BRM.40.3.879

Pressley, M., Forrest-Pressley, D. L., Elliott-Faust, D., & Miller, G. (1985). Children's use of cognitive strategies, how to teach strategies, and what to do if they can't be taught. In M. Pressley & C. J. Brainerd (Eds.), *Cognitive learning and memory in children* (pp. 1–47). Springer. https://doi.org/10.1007/978-1-4613-9544-7_1

Primi, R., Ferrão, M. E., & Almeida, L. S. (2010). Fluid intelligence as a predictor of learning: A longitudinal multilevel approach applied to math. *Learning and Individual Differences*, *20*(5), 446–451. https://doi.org/10.1016/j.lindif.2010.05.001

R Core Team. (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Raven, J. (2000). The Raven's Progressive Matrices: Change and stability over culture and time. *Cognitive Psychology*, *41*(1), 1–48. https://doi.org/10.1006/cogp.1999.0735

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

Raven, J. C. (1941). Standardisation of progressive matrices, 1938. *British Journal of Medical Psychology*, *19*(1), 137–150. https://doi.org/10.1111/j.2044-8341.1941.tb00316.x

Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment*, *28*(3), 164–171. https://doi.org/10.1027/1015-5759/a000123

Ricco, R. B. (2015). The development of reasoning. In L. S. Liben, U. Müller, & R. M. Lerner (Eds.), *Handbook of child psychology and developmental science: Cognitive processes* (Vol. 2, 7th ed., pp. 519–570). John Wiley & Sons. https://doi.org/10.1002/9781118963418.childpsy213

Rivollier, G., Quinton, J.-C., Gonthier, C., & Smeding, A. (2021). Looking with the (computer) mouse: How to unveil problem-solving strategies in matrix reasoning without eye-tracking. *Behavior Research Methods*, *53*(3), 1081–1096. https://doi.org/10.3758/s13428-020-01484-3

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Rozencwajg, P., Cherfi, M., Ferrandez, A. M., Lautrey, J., Lemoine, C., & Loarer, E. (2005). Age related differences in the strategies used by middle aged adults to solve a block design task. *The International Journal of Aging and Human Development*, *60*(2), 159–182. https://doi.org/10.2190/H0AR-68HR-RRPE-LRBH

Rozencwajg, P., & Corroyer, D. (2001). Strategy development in a block design task. *Intelligence*, *30*(1), 1–25. https://doi.org/10.1016/S0160-2896(01)00063-0

Schneider, W. (2008). The development of metacognitive knowledge in children and adolescents: Major trends and implications for education. *Mind, Brain, and Education*, *2*(3), 114–121. https://doi.org/10.1111/j.1751-228X.2008.00041.x

Siegler, R. S. (2000). The rebirth of children's learning. *Child Development*, *71*(1), 26–35. https://doi.org/10.1111/1467-8624.00115

Siegler, R. S. (2016). Continuity and change in the field of cognitive development and in the perspectives of one cognitive developmentalist. *Child Development Perspectives*, *10*(2), 128–133. https://doi.org/10.1111/cdep.12173

Siegler, R. S., Adolph, K. E., & Lemaire, P. (1996). Strategy choices across the life span. In L. R. Reder (Ed.), *Implicit memory and metacognition* (pp. 79–121). Erlbaum.

Siegler, R. S., & Svetina, M. (2002). A microgenetic/cross-sectional study of matrix completion: Comparing short-term and long-term change. *Child Development*, *73*(3), 793–809. https://doi.org/10.1111/1467-8624.00439

Simmering, V. R., & Perone, S. (2013). Working memory capacity as a dynamic process. *Frontiers in Psychology*, *3*, Article 567. https://doi.org/10.3389/fpsyg.2012.00567

Smedslund, J. (1964). Concrete reasoning: A study of intellectual development. *Monographs of the Society for Research in Child Development*, *29*(2), 1–39. https://doi.org/10.2307/1165716

Snow, R. E. (1978). Eye fixation and strategy analyses of individual differences in cognitive aptitudes. In A. M. Lesgold, J. W. Pellegrino, S. D. Fokkema, & R. Glaser (Eds.), *Cognitive psychology and instruction* (pp. 299–308). Plenum Press.

Snow, R. E. (1980). Aptitude processes. In R. E. Snow, P.-A. Federico, & W. E. Montague (Eds.), *Aptitude, learning, and instruction: Cognitive process analyses of aptitude* (Vol. 1, pp. 27–63). 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

Styles, I. (2008). Linking psychometric and cognitive-developmental frameworks for thinking about intellectual functioning. In J. Raven & J. Raven (Eds.), *Uses and abuses of intelligence: Studies advancing Spearman and Raven's quest for non-arbitrary metrics* (pp. 69–98). Royal Fireworks Press; Competency Motivation Project; EDGE 2000; RTS Romanian Psychological Testing Services SRL.

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

Thibaut, J.-P., French, R., & Vezneva, M. (2010a). Cognitive load and semantic analogies: Searching semantic space. *Psychonomic Bulletin & Review*, *17*(4), 569–574. https://doi.org/10.3758/PBR.17.4.569

Thibaut, J.-P., French, R., & Vezneva, M. (2010b). The development of analogy making in children: Cognitive load and executive functions. *Journal of Experimental Child Psychology*, *106*(1), 1–19. https://doi.org/10.1016/j.jecp.2010.01.001

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505. https://doi.org/10.3758/BF03192720

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

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

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

Whitebread, D. (1996). The development of children's strategies on an inductive reasoning task. *British Journal of Educational Psychology*, *66*(1), 1–21. https://doi.org/10.1111/j.2044-8279.1996.tb01172.x

Wood, S. N. (2017). *Generalized additive models: An introduction with R* (2nd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781315370279

Received January 25, 2023
Revision received October 1, 2023
Accepted October 7, 2023 ■