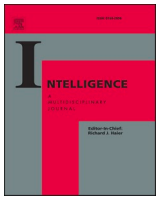
Intelligence 116 (2026) 102020

Contents lists available at ScienceDirect

Intelligence

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



# Rethinking rule diversity in figural matrices: A log file analysis on the role of task switching and implications from a validation study

Dominik Weber [a,*], Stella Jelen [a], Frank M. Spinath [a], Florian Krieger [b], Nicolas Becker [a], Marco Koch [a]

[a] Department of Individual Differences & Psychodiagnostics, Saarland University, Campus A1 3, D-66123 Saarbrücken, Germany
[b] Department of Methods of Educational Research, TU Dortmund University, Emil-Figge-Straße 50, D-44227 Dortmund, Germany

ARTICLE INFO

*Keywords:*
Figural matrices
Reasoning
Fluid intelligence
Task switching
Cognitive flexibility
Log files
Process data
Response times
Construct validity

ABSTRACT

Figural matrices typically consist of multiple distinct logical rules, requiring test-takers to disengage from one rule before applying the next. Prior studies have consistently shown moderate associations between matrix performance and task-switching ability. However, these findings are largely based on correlational data. The present two-study article aimed to (a) determine, on a theoretical level, whether task-switching ability is functionally involved in matrix processing, and (b) assess, from a diagnostic perspective, whether relaxing the constraint of distinct rules within a single matrix threatens psychometric validity. To this end, we manipulated matrices to include both distinct-rule and identical-rule transitions, enabling experimental within-subject comparisons of matrix processing in both conditions based on log file analyses. In study 1 ($N = 209$), task-switching ability exerted a functional influence only during distinct-rule transitions. However, the correlation between task-switching ability and matrix performance remained comparably strong even during identical-rule transitions. This dual pattern supports both a switch-dependency hypothesis (i.e., that task-switching is functionally involved in matrix processing) and a shared-resource hypothesis (i.e., that task-switching and matrix processing draw on a common cognitive resource). In study 2 ($N = 258$), we evaluated the convergent validity of the newly designed mixed-rule format against the traditional distinct-rule format. Test scores were highly correlated ($r = 0.87$), and test characteristics (e.g., reliability, IRT and TIF parameters) and external validity were very similar. Taken together, these findings suggest that although task-switching demands can vary depending on matrix design, relaxing rule constraints does not compromise psychometric validity. This flexibility in item development may be particularly useful in large-scale assessments or student selection tests that require continuous item renewal.

## 1. Introduction

Figural matrices are widely used as an economical and valid proxy for reasoning ability (Marshalek, Lohman, & Snow, 1983; Jensen, 1998), which constitutes a core component of general intelligence (Carroll, 1993; McGrew, 2009). Each item typically consists of multiple logical rules (e.g., addition, subtraction, intersection) that test-takers must identify and apply in order to solve a matrix. Traditionally, these rules are distinct within a single matrix (i.e., each rule appears only once per item). This design requires test-takers to disengage from one rule after solving it and then to shift to a different and unrelated rule, potentially placing demands on cognitive flexibility. Previous research has consistently shown moderate correlations between figural matrix performance and task-switching paradigms as an operationalization of cognitive flexibility (e.g., Li, Li, Stoet, & Lages, 2019; Salthouse, Fristoe, McGuthry, & Hambrick, 1998; Yehene & Meiran, 2007). However, it remains unclear whether this association reflects a causal relationship (implying that task-switching ability is necessary for solving figural matrices) or merely reflects shared variance (e.g., due to a shared underlying cognitive resource). The aims of this article were (a) to determine the nature of this association, and building on these theoretical insights (b) to examine whether allowing for identical rules within a single matrix (rather than requiring distinct rules only) compromises psychometric quality and external validity of figural matrix tests.

* Corresponding author at: Saarland University, Department of Individual Differences & Psychodiagnostics, Campus A1 3/3.05, Germany.
*E-mail address:* dominik.weber@uni-saarland.de (D. Weber).

https://doi.org/10.1016/j.intell.2026.102020
Received 23 December 2025; Received in revised form 23 March 2026; Accepted 21 April 2026
Available online 28 April 2026

### 1.1. Figural matrices in intelligence assessment

Figural matrices typically consist of 3 × 3 grids containing abstract symbols (e.g., Formann, Waldherr, & Piswanger, 2011; Pallentin, Danner, & Rummel, 2023) that follow specific logical rules, such as addition, rotation, or intersection. The last cell is empty and must be completed by inferring the underlying rules (Fig. 1A). Most matrices do not only contain one, but multiple rules, each of which is distinct, requiring test-takers to identify and integrate several unrelated transformations. While traditional matrix tests provide predefined response options (e.g., Raven, 1962), in the past decade construction-based matrix formats have been introduced (Becker & Spinath, 2014; Koch, Spinath, Greiff, & Becker, 2022; Krieger, Becker, Greiff, & Spinath, 2022). These items require test-takers to actively assemble the correct solution using a construction kit containing all possible symbols, thereby minimizing guessing probability. Moreover, construction-based matrices prevent participants from arriving at the correct solution merely by successively eliminating response options without comprehending the underlying logical rules. Compared to such response-elimination strategies, constructive matching strategies (i.e., the mental construction of the solution to an item), which is fostered by the construction-based format, have been shown to result in higher validity (e.g., Arendasy & Sommer, 2013; Vigneau, Caissie, & Bors, 2006). Overall, construction-based matrix tests have demonstrated high internal consistency (α = 0.87–0.96), strong construct validity and consistent external validity with educational outcomes such as GPA or level of education ($r$ = 0.13–0.32; Becker et al., 2016; Becker & Spinath, 2014; Krieger et al., 2022; Weber et al., 2023). These psychometric qualities have contributed to the increasing popularity of figural matrix tests in modern intelligence diagnostics. For example, the main German student selection test for psychology (BaPsy; e.g., Schulz-Hardt, 2025) includes a figural matrix scale consisting of 20 items, accounting for approximately 16% of the total items in the assessment. Additionally, substantial convergent validity has been observed between construction-based figural matrices and several subtests of the German student selection tests for medicine (Levacher et al., 2023).

### 1.2. Log file analyses as a window into figural matrix processes

Research has shown that successful matrix solving depends not only on rule induction as a reasoning function, but also on executive processes such as goal management (Carpenter, Just, & Shell, 1990). Goal management refers to the ability to decompose a complex task (e.g., solving an entire matrix) into subgoals (e.g., solving single matrix rules), to solve these subgoals structurally, and to integrate the partial solutions into a total solution (e.g., Embretson, 1998; Krieger, Zimmer, Greiff, Spinath, & Becker, 2019; Loesche, Wiley, & Hasselhorn, 2015). When implemented in computer-based assessments, construction-based matrices generate detailed log file data that track test-takers' actions (e.g., clicks) along with corresponding timestamps. Log file analyses have become a powerful methodological approach in contemporary digital assessment (e.g., Chen, Liu, & Mao, 2024; Nicolay, Krieger, & Greiff, 2023), since they allow fine-grained analyses of participants' problem-solving strategies beyond traditional scoring (Stadler, Hofer, & Greiff, 2020).

Recently, two key indicators of structured matrix processing in terms of goal management have emerged from log file analyses: First, rule jumps, defined as premature switching between rules before fully solving the current one. In this regard, Weber et al. (2023) found that fewer rule jumps were strongly associated with better matrix-test performance ($r$ = −0.54). Second, temporal characteristics of matrix processing, namely interrule times, defined as the latency between processing two consecutive rules (Weber, Koch, Spinath, Krieger, & Becker, 2025). With respect to these temporal characteristics of matrix processing, previous research has revealed interindividual differences in the relationship between time on task (ToT) and test performance: In lower ability ranges, longer ToT was associated with better performance, while in higher ability ranges, the relationship was reversed (Becker, Schmitz, Falk, et al., 2016; Goldhammer, Naumann, & Greiff, 2015; Krämer, Koch, Levacher, & Schmitz, 2023). Analyses of log file data suggest that this pattern primarily originates from differences in interrule times, which serve as indicators of different matrix processing strategies (Weber et al., 2025). Based on interrule times, participants could be clustered into structured (rule-by-rule) and unstructured solvers (characterized by frequent rule jumps). Among structured solvers, shorter interrule times were associated with higher test performance ($r$ = −0.51), and the temporal parameters derived from log files accounted for an incremental 24.30% of the performance variance beyond traditional ToT.

Although a substantial body of evidence has accumulated regarding

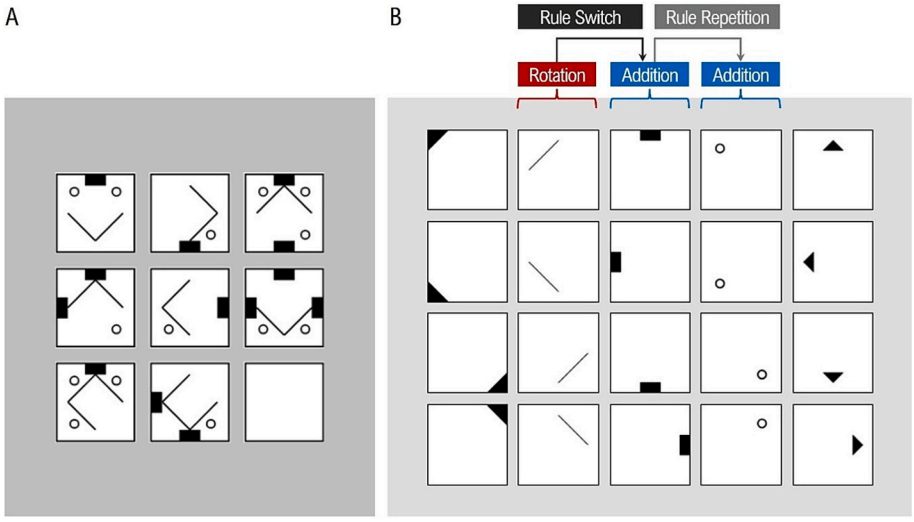



**Fig. 1.** Illustration of the principles of figural matrices. **A.** The item stem consists of a 3 × 3 grid containing different symbols that follow specific logical rules (e.g., the black rectangles in the first two cells of a row add up to the rectangles in the third cell). The last cell is empty and must be completed by applying the inferred logical rules. **B.** In the construction-based item format, participants use a construction kit to build the solution. In this example of the re-designed item format, the matrix includes three symbol groups that follow two different rules. If participants process first the line symbol group (rotation) and then the rectangle symbol group (addition), this constitutes a distinct-rule transition. If they subsequently process the circle symbol group (addition), this constitutes an identical-rule transition.

figural matrix processing, relatively little is known about the relations to more basic cognitive processes. In fact, an ongoing debate spanning more than 30 years has addressed whether general cognitive processes are essential for solving figural matrices (e.g., Frischkorn & Oberauer, 2021; Unsworth & Engle, 2005), or whether specific task demands of figural matrices require specific cognitive processes (e.g., Krieger et al., 2019).

### 1.3. Task-switching ability and its relationship to matrix-test performance

Besides structuredness of matrix processing, another possible explanation for interindividual differences in temporal variables, particularly in interrule times between two logical rules, is variability in task-switching ability. Task-switching ability is widely regarded as a key indicator of cognitive flexibility (e.g., Kiesel et al., 2010; Koch, Poljac, Müller, & Kiesel, 2018) and well established in the literature as an executive function alongside updating (of working-memory representations) and inhibition (Miyake et al., 2000). In experimental research, various types of paradigms have been employed to assess task switching ability, in each case contrasting situations with low vs. high demands on cognitive flexibility. A particularly prominent paradigm is the so-called task-cueing procedure (e.g., Kiesel et al., 2010), in which a cue signals which of at least two tasks must be applied on an upcoming stimulus. Trials can differ in stimulus-task mappings: they may be univalent (i.e., only one task is applicable) or bivalent/multivalent (i.e., both/multiple tasks can be applied to the same stimulus). For example, when participants must decide whether digits are odd vs. even or smaller vs. larger than 5 (e.g., Sudevan & Taylor, 1987), the same digit can be used in both tasks, rendering it a bivalent stimulus. In contrast, if task A involves digits (e.g., odd vs. even) and task B involves letters (e.g., vowel vs. consonant), stimuli are univalent. Bivalent stimuli generally elicit longer response times due to increased task interference (e.g., Rogers & Monsell, 1995). When tasks are presented in randomized order, trials can be classified as either switch (AB) or repeat (AA) trials. Responses are typically slower in switch trials, and the difference in mean response time between switch and repeat trials is commonly used as an index of task-switching ability, referred to as switch costs (e.g., Kiesel et al., 2010). Notably, while overall response times can be reduced by lengthening the cue-stimulus interval, switch costs themselves often remain unaffected (e.g., Altmann, 2004; Koch, 2001, 2005), which suggests that switch costs reflect more stable cognitive demands. Several studies have investigated what drives the difference in response times between switch and repeat trials: One central explanatory mechanism is task-set inertia (Allport & Wylie, 2000; Schmitz & Krämer, 2023), which refers to proactive interference from previously activated task set. In other words, task-set inertia describes the cognitive effort to inhibit the prior task and to engage fully with the current one (e.g., Allport, Styles, & Hsieh, 1994).

Previous research has consistently demonstrated moderate associations between task-switching ability and performance on figural matrices. For example, Salthouse et al. (1998) reported correlations between task-switching ability and performance in Raven's Matrices ranging from $r = -0.31$ to $r = -0.46$ (note that smaller switch costs indicate higher ability). Similarly, Yehene and Meiran (2007) found a correlation of $r = -0.38$ between task-switching ability and a general intelligence score, based on Raven's Matrices, a vocabulary test, and a perceptual speed test. More recently, Li et al. (2019) reported that participants with high task-switching ability outperformed those with lower abilities on Raven's Advanced Progressive Matrices ($d = 1.19$). However, while the relationship between task-switching ability and matrix-test performance is well-established, the underlying nature of this association remains unclear. Prior studies have typically relied on correlational instead of experimental designs, leaving open the question of whether task-switching ability is functionally involved in matrix processing or not.

### 1.4. The present research

The goals of this research were (a) to investigate, on a theoretical level, whether task-switching ability plays a necessary role in figural matrix processing, and (b) to derive, on a diagnostic level, implications for item design building on these theoretical insights and on a psychometric validation.

The consistently reported correlations raise the question of what drives the relationship between task-switching ability and figural matrices (RQ 1). One possibility is that task-switching ability is functionally required for processing figural matrices (*switch-dependency hypothesis*). This hypothesis appears plausible, as a figural matrix typically contains several distinct rules, requiring test-takers to disengage from the most recently applied rule and shift to the next one. Although rule transitions in figural matrices and task switches in experimental paradigms differ in several aspects such as timing control and task complexity, both contexts involve cognitive reconfiguration (i.e., disengaging from a prior rule or task and engaging with a new one). This conceptual parallel allows for examining the functional role of cognitive flexibility in a highly naturalistic way beyond traditional task-switching paradigms. Building on this, if the switch-dependency hypothesis is accurate, higher task-switching ability should be associated with shorter interrule times (i.e., with shorter latencies between processing two consecutive matrix rules) which are linked to better matrix performance. In other words, interrule times should partially mediate the relationship between task-switching ability and matrix-test performance.

An alternative explanation for the relationship between task-switching ability and matrix-test performance is that both measures draw on a common underlying cognitive resource (e.g., executive functions or attention control). According to this *shared-resource hypothesis*, cognitive capacity is allocated depending on current task demands. In this case, the association between task-switching ability and matrix performance would reflect a correlational overlap rather than a functional dependency and would not be expected to impact interrule times or matrix performance directly.

(Indirect) support for the switch-dependency hypothesis comes from research examining the association between performance on figural matrices and working memory capacity (WMC): Shipstead and Engle (2013) provided participants with a battery of fluid intelligence tests (including Raven's Matrices; Raven, 1962), a WMC task, and a visual arrays task. In the latter task, a pattern of symbols was presented, followed by a blank-screen interval and a second pattern, to which participants responded with a same-different judgment. The duration of the interval between the patterns was manipulated to allow for varying degrees of disengagement, which serves to reduce memory interference. The results showed that particularly participants with higher fluid intelligence benefited from enlarging the interval, and that the correlation between fluid intelligence and the performance on the visual arrays task increased with increasing interval length. In contrast, the correlation between WMC and the performance on the visual arrays task remained stable across interval length. This finding was interpreted as evidence that disengagement is more strongly implicated in fluid intelligence than in WMC (analogously to its role in task-switching ability).

WMC appears to be particularly strongly associated with the performance on figural matrix tasks when these tasks place demands on secondary memory: To test this assumption, Harrison, Shipstead, and Engle (2015) constructed item pairs of Raven's matrices that shared identical rule combinations within pairs but differed in rule combinations across pairs. Matrices with identical and distinct combinations were presented in alternation, yielding two corresponding performance scores. Performance on a WMC task correlated more strongly with the score based on the identical combinations than with the score based on the distinct rules (but see Wiley, Jarosz, Cushen, & Colflesh, 2011, who found a opposite pattern of effects). These results indicate that the correlation between figural matrices and WMC is not driven by disengagement processes. Although task-switching ability and WMC are

separable constructs, both are traditionally subsumed under executive functions (Miyake et al., 2000) and substantially correlated (e.g., Himi, Bühner, Schwaighofer, Klapetek, & Hilbert, 2019; Oberauer, Süß, Wilhelm, & Wittman, 2003). Accordingly, the findings reported by Harrison et al. may be interpreted as cautious support for the shared-resource hypothesis.

Later, Shipstead, Harrison, and Engle (2016) integrated these findings into a broader theoretical model. The core construct of this model is top-down executive attention, which is expressed through two cognitive functions: maintenance, which protects task-relevant information from interference, and disengagement, which facilitates the removal of no-longer relevant information. The authors propose that WMC tasks primarily (but not exclusively) place demands on maintenance, whereas fluid intelligence tests such as figural matrices place greater demands on disengagement. Consequently, this model accommodates both the switch-dependency and shared-resource hypothesis would be reflected.

Traditional matrix designs make it difficult to determine experimentally whether the link between task-switching ability and matrix processing is causal or merely correlational, as it typically includes only distinct rules within an item. Despite their valuable contribution to understanding the cognitive processes underlying figural matrices, the research of Harrison et al. focused exclusively on transitions between entire rule combinations (i.e., between total items) rather than on transitions between single rules. However, transitions at level of single rules are particularly informative for elucidating elemental cognitive processes, whereas analyzing transitions between total items could mask these elemental processes. To address these limitations, we developed a new matrix design that allows for both distinct and identical rules within a single item. This variation constitutes an experimental manipulation, enabling within-subject comparisons between the processing behavior and performance on distinct-rule transitions versus identical (i.e., repeated)-rules transitions. By directly contrasting these two conditions, this design offers a more fine-grained understanding of the cognitive mechanisms underlying in matrix performance and its link to task-switching ability.

Building on the findings related to RQ 1 on the role of task-switching ability, we aimed to draw practical inferences for matrix design in test development and diagnostic applications. Specifically, we examined whether it is reasonable to rethink the principle of rule diversity in figural matrices. Traditional matrices typically include only distinct logical rules per item, which limits the number of possible rule combinations and reduces the degrees of freedom in rule sequences. As a result, once a subset of rule types has been identified, solving the remaining ones may become easier, because test-takers can implicitly assume that previously applied rules will not reappear within the same item. For item development, especially in student selection tests or large-scale assessments, relaxing such constructional constraints could therefore be beneficial. This raises the question of whether allowing identical rules alongside distinct rules within a single matrix would compromise psychometric quality and external validity (RQ 2). To answer these two intertwined research questions, we conducted the following two studies.

## 2. Study 1

In study 1, we addressed RQ 1, which concerns the nature of the relationship between task-switching ability and matrix-test performance. Specifically, we tested whether task-switching ability is functionally necessary for successful matrix solving (switch-dependency hypothesis), or whether the frequently observed correlation between both abilities is due to a shared underlying cognitive resource without a causal link (shared-resource hypothesis). To analyze this, we tested four preregistered hypotheses (H1–H4; OSF: https://osf.io/fd6qg/?view_only=fae71735cc9640c780ca2dc79b102929). The expected direction and magnitude of effects were informed by prior research:

H1. Following Weber et al. (2025), we expected to identify two participant clusters based on interrule times as an indicator for structured matrix processing and matrix performance. We hypothesized that one cluster would show (a) longer interrule times (reflecting higher elaboration of a mental solution) and (b) fewer rule jumps, both indicating structured processing. Within this structured cluster, we expected (c) a strong negative correlation between the interrule times derived from log data and matrix performance that explains incremental variance beyond traditional ToT. Based on the findings of Weber et al. (2025), we expected an incremental variance explanation of approximately $\Delta R^2 = 24.30\%$. This replicative hypothesis aimed to provide by the following analyses a more nuanced understanding of the relationship between task-switching ability and matrix processing for subgroups with different processing strategies.

H2. We expected longer interrule times when participants processed distinct rules of a matrix (e.g., addition followed by intersection) compared to identical rules (e.g., intersection followed by intersection). We further hypothesized that switch costs from a task-switching paradigm would correlate more strongly with interrule times between distinct rules than with those between identical rules.

H3. We expected poorer matrix performance when sequential matrix rules were distinct rather than identical. Again, we assumed that switch costs from a task-switching paradigm would show stronger associations with performance on distinct rules than on identical rules. Salthouse et al. (1998) and Yehene and Meiran (2007) reported correlations ranging from $r = -0.31$ to $-0.46$ between task-switching ability and matrix performance. Because these associations were observed in matrix tests in which each item contained distinct rules, correlations of comparable magnitude were expected for performance on distinct rules in the present study. In contrast, lower correlations were expected for performance on identical-rule items.

H4. We expected that interrule times would mediate the correlation between switch costs and matrix performance more strongly when participants were required to process distinct rules than when they processed identical rules.

To prepare for RQ 2 of whether it is reasonable to allow for identical rules within a single matrix, we examined as post-hoc analysis the convergent validity between performance on distinct rules and on identical rules. Additionally, we determined the external validity of both performance measures with participant's GPA and level of education. Based on previous findings on the association between construction-based matrix tests and educational variables, correlations in the range of $r = 0.13$ to 0.32 were expected (Becker, Schmitz, Falk, et al., 2016; Becker & Spinath, 2014; Krieger et al., 2022; Weber et al., 2023).

### 2.1. Methods

#### 2.1.1. Sample and exclusion criteria

We initially recruited $N = 302$ participants between March and June 2025 who participated in all parts of the study. Participants were acquired at the local university, in the upper secondary education (academic track) and on the survey platform SurveyCircle (SurveyCircle, 2025) available in Germany, Switzerland and Austria. As preregistered, we excluded 34 participants who, in a post-study questionnaire, reported that they had guessed while processing the figural matrix test. Another 46 participants were excluded due to implausibly fast response times on the matrix task. Implausibly fast response times were defined as values below 23.21 s per matrix, corresponding to the mean response time minus two standard deviations reported in the study by Weber et al. (2025). Additionally, 6 participants showed extreme values (defined as $>3 \times$ interquartile range regarding reaction time) in the task-switching paradigm, and for further 7 participants, either the independent or

dependent variables could not be computed due to missing data. The final sample consisted of $N = 209$ participants (69.38% female, 28.71% male, 1.91% non-binary), with a mean age of $M = 27.27$ years ($SD = 11.00$). The sample was heterogeneous in terms of educational background: Appendix A presents the distribution of participants according to the International Standard Classification of Education Fields of Education and Training (ISCED-F) provided by UNESCO (UNESCO, 2015). Participants from the local university received course credits for their participation.

### 2.1.2. Test procedure and materials

*2.1.2.1. Figural matrices.* Participants were assessed with 20 computer-based items adapted from the Open Matrices Item Bank (OMIB; Koch et al., 2022). These items are construction-based, with the consequence that participants must build their response using a predefined construction-kit containing all possible symbols. Five items included three rules, ten items included four rules, and five items included five rules. Importantly, to implement experimental manipulation, items were modified to contain both distinct and identical rules within a single matrix. This allowed for a within-subject design, enabling the assessment of each participant's processing behavior and performance under both of two conditions (Fig. 1B): a condition with traditional distinct-rule transitions as well as a condition with identical-rule transitions as a baseline. Participants had a maximum of 75 s to solve each item. Although some traditional paper-and-pencil based matrix tests such as the Standard Progressive Matrices (Raven, Court, & Raven, 1996) do not impose a time limit according to the test manual, item-level time limits have proven useful and economic in computerized administrations of matrix tests (e.g., Domnick, Zimmer, Becker, & Spinath, 2017; Krieger et al., 2022). In the present studies, the time limit was chosen based on the empirically observed mean response time of $M = 51.79$ s per item by Weber et al. (2025).

*2.1.2.2. Task-switching paradigm.* In addition to the matrix test, participants completed a task-switching paradigm. To closely match the structure of the matrix items, we used mixed blocks of bivalent stimuli (i.e., stimuli in which both of two tasks within a block could potentially apply). The paradigm comprised four blocks of 32 bivalent trials each: two numeric and two figural blocks. In the numeric blocks, digits from 1 to 9 were presented, and participants had to decide whether the digit was (a) odd vs. even, or (b) smaller vs. larger than 5. In the figural blocks, geometric shapes were presented, and participants had to decide whether the figure was (a) round vs. angular, or (b) black vs. white. Each trial began with a fixation cross presented for 1000 ms, followed by a 200 ms mask. Then a task cue was displayed for 1000 ms, indicating the relevant task. Following the cue, the target stimulus appeared and remained on screen until the participant responded. Participants responded by pressing the left or right arrow key on the keyboard (Fig. 2). No feedback was provided. Prior to each block, an instruction and an example trial were displayed. Stimuli were presented within an 8 × 8 cm (3.15 × 3.15 in) display box using a 33 pt. font. The stimulus sequences were randomly generated with *R* (R Core Team, 2025) ensuring as many switch (AB) as repeat (AA) trials. The order of tasks (matrix test first vs. task-switching paradigm first) was randomized across participants. After processing the cognitive tests, participants were asked to complete a demographic questionnaire, which included their GPA and level of education. The assessment was programmed using the JavaScript interface in Unipark (Tivian XI GmbH, 2025).

### 2.1.3. Scoring and derived measures

To determine overall matrix performance, we applied the partial solution procedure (Weber et al., 2023), determining the number of matrix rules correctly solved by each participant across the entire test. To differentiate performance on distinct vs. identical rules, we computed two separate indices based on the following preregistered formulas:

$$\textit{switch index} = \frac{\textit{number of rules solved after rule switch}}{\textit{number of rules attempted after rule switch}} \times \frac{\textit{total number of rules solved}}{\textit{total number of rules}} \tag{1}$$

$$\textit{repetition index} = \frac{\textit{number of rules solved after rule repetition}}{\textit{number of rules attempted after rule repetition}} \times \frac{\textit{total number of rules solved}}{\textit{total number of rules}} \tag{2}$$

The first quotient in each index represents the accuracy after distinct-rule transitions (switch index) or identical-rule transitions (repetition index), respectively. Since traditional matrix performance metrics do not rely solely on accuracy (items solved / items attempted), but also reflect processing efficiency (sum score), we multiplied the accuracy term by the participant's overall solution rate to obtain a weighted score which considers both accuracy and efficiency. Consequently, the resulting indices ranged from 0 to 1, where values close to 1 reflect high accuracy *and* high efficiency. From log file data, we additionally computed the mean interrule time (i.e., the latency between two consecutive rules), separately for distinct-rule and identical-rule transitions. Furthermore, we calculated the number of rule jumps as an indicator of structured matrix processing (Weber et al., 2023).

From the task-switching paradigm, we computed switch costs as the mean difference in response time between switch trials and repetition trials within each block. To account for differences in baseline reaction times between blocks, we standardized the switch costs per block and then computed the mean switch costs across the four blocks.

### 2.1.4. Statistical analysis

To classify participants based on their matrix processing strategy, we conducted a cluster analysis using participants' interrule times and partial solution score. Specifically, we performed hierarchical agglomerative clustering using Ward's method with Euclidean distance, applying 1000 bootstrap iterations (cf. Weber et al., 2025). To validate the clustering in terms of processing strategy, we compared the resulting clusters with respect to mean interrule times and mean number of rule jumps. Additionally, we conducted a hierarchical regression analysis to test whether, within the structured cluster, interrule times predicted test performance beyond the ToT.

To evaluate the effectiveness of our experimental manipulation, we compared the interrule times between distinct-rule and identical-rule

transitions. Within each cluster (structured vs. unstructured), we computed the correlation between switch costs and interrule times separately for distinct-rule and identical-rule transitions. To assess whether the manipulation had an influence on matrix-test performance, we tested the difference between the switch index and repetition index for significance. Furthermore, we calculated the correlations between switch costs and both performance indices and, in case of two significant correlations, compared them using a *t*-test for dependent correlations. Crucially, we examined in both clusters whether interrule times mediated the relationship between switch costs and the respective performance index. These mediation analyses served to clarify whether task-switching ability plays a functional role in matrix processing.

Finally, to draw preliminary diagnostic inferences and to prepare for study 2, we tested whether it is psychometrically meaningful to differentiate matrix performance on identical rules from the traditional performance on distinct rules. To this end, we conducted two confirmatory factor analyses (CFA): (1) a single-factor model, where the performance on both the distinct-rules and the identical-rule transitions loaded on one common factor; (2) a two-factor model, where the two performances loaded on separate factors. For both models, we constructed three parcels each by alternating item assignment across parcels (i.e., parcel 1 = item 1, 4, 7, …; parcel 2 = item 2, 5, 8, …; parcel 3 = item 3, 6, 9, …). Item parceling has been described as a reasonable technique, particularly when applied to unidimensional scales (e.g., Bandalos & Finney, 2001; Little, Cunningham, Shahar, & Widaman, 2002) such as figural matrices. We compared model fits to evaluate whether extracting two factors is reasonable. In addition, we examined the correlations of both indices with external criteria, namely participants' GPA and level of education.

A total of 93 participants from the original sample were excluded from the main analyses. Although these exclusions were based on preregistered criteria, we conducted additional analyses due to the comparatively large number of excluded cases to examine whether the core findings would remain robust when all participants were retained. To rule out potential selection bias, we therefore report robustness analyses based on the full sample ($n = 302$) alongside the results obtained from the data cleaned according to predefined quality criteria. In addition, we examined whether the relatively large number of participants who reported guessing could be attributed to the imposed per-item time limit. Finally, since the cluster structure forms the basis for the analyses of subsamples with different processing strategies, we conducted robustness analyses using alternative clustering specifications and examined whether participants were assigned to the same clusters across methods. To this end, we varied clustering specifications in three steps of increasing deviation from the original method: the distance metric (Ward's method with Mahalanobis distance), the linkage method (complete linkage with Euclidean distance), and the clustering algorithm (k-means clustering).

We used the *R* (R Core Team, 2025) packages *dplyr* (Wickham et al., 2023) and *stringr* (Wickham, 2023) for data preparation, the packages *effectsize* (Ben-Shachar, Lüdecke, & Makowski, 2020), *lavaan* (Rosseel, 2012), *lm.beta* (Behrendt, 2023), and *psych* (Revelle, 2025) for data analysis, and the package *factoextra* (Kassambara & Mundt, 2020) for visualization. The hypotheses and analysis plan were preregistered on OSF: https://osf.io/fd6qg/?view_only=fae71735cc9640c780ca2dc79b102929.

### 2.2. Results

#### 2.2.1. Descriptive statistics

On average, participants correctly solved $M = 52.76$ of 80 matrix rules ($SD = 19.78$). The switch index as a performance measure on distinct-rule transitions was $M = 0.38$ ($SD = 0.20$), and the repetition index as a performance measure on identical-rule transitions was $M = 0.37$ ($SD = 0.19$). Participants spent an average of $M = 56.08$ s ($SD = 12.48$) per matrix. The mean interrule time (i.e., the latency between the processing of two rules) was $M = 11.80$ s ($SD = 3.24$). Specifically, participants spent $M = 12.30$ s ($SD = 3.53$) between distinct rules and $M = 11.18$ s ($SD = 3.76$) between identical rules. In the task-switching paradigm, participants showed mean switch costs of $M = 98.25$ ms ($SD = 140.82$). internal consistencies for the key variables were high: $\alpha_{score} = 0.96$, $\alpha_{switch} = 0.88$, $\alpha_{repetition} = 0.80$, $\alpha_{interrule} = 0.92$. Detailed item-level statistics for the matrix test are provided in Appendix C.

#### 2.2.2. Replication of the role of the Interrule times

The cluster analysis based on interrule times and matrix performance resulted in two plausible clusters, replicating the pattern reported by Weber et al. (2025); Fig. 3). One cluster ($n = 172$) invested significantly more time between the rules than the other one ($n = 37$), $t(50.05) = 11.40$, $p < .001$, $d = 2.18$ (95% CI [1.76, 2.59]), and showed fewer rule jumps, $t(39.56) = -3.06$, $p = .002$, $d = -0.86$ (95% CI [−1.22, −0.49]), indicating a more structured matrix processing strategy. Among participants in this cluster, interrule times strongly predicted matrix-test performance (i.e., the faster participants transitioned between rules, the better their overall performance), $\beta = -0.61$, 95% CI [−0.69, −0.52], $p < .001$. In a hierarchical regression analysis, interrule times explained an incremental $\Delta R^2 = 32.90\%$ of the variance in matrix performance beyond ToT, which was a significant model improvement: $F(1,$

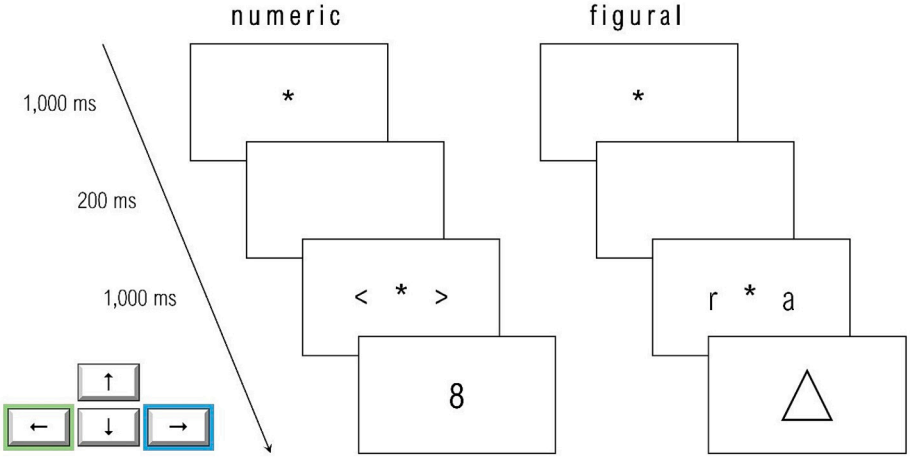



**Fig. 2.** Schematic of the task-switching paradigms. Participants completed two types of task-cueing procedures: one numeric and one figural. Each trial began with a fixation cross presented for 1000 ms, followed by a 200 ms mask. A task cue was then shown for 1000 ms, indicating which of two tasks to perform on the upcoming stimulus. Participants responded by pressing the left or right key on the keyboard. r = round, a = angular.

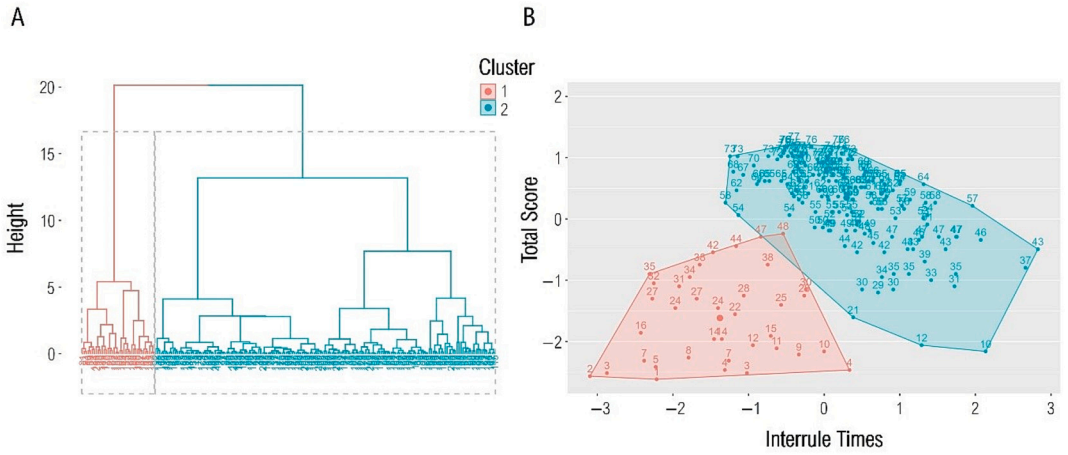



**Fig. 3.** Results of the cluster analysis. A. The dendrogram indicates two main branches of acceptable height, suggesting a two-cluster solution. B. The scatterplot shows the participants assigned to either a cluster with structured ($n = 172$) or unstructured matrix processing ($n = 37$) based on their interrule times and total score. Variables were standardized for clustering, raw (rounded) interrule times are displayed above each data point for reference.

169) = 89.89, $p < .001$, $R^2 = 38.16\%$.

*2.2.3. Is task-switching ability functionally involved in figural matrix processing?*

As hypothesized, participants took significantly more time for distinct-rule transitions compared to identical-rule transitions: $\Delta M = 1.12$ s per transition, $t(208) = 6.80$, $p < .001$, $d = 0.33$ (95% CI [0.13, 0.52]). However, performance was not significantly better on identical rules than on distinct rules: $t(208) = 0.29$, $p = .998$. Across the full sample, switch costs from the task-switching paradigm were not significantly correlated with interrule times, neither for distinct-rule transitions ($r = 0.07$, $p = .304$) nor for identical-rule transitions ($r = 0.08$, $p = .232$). However, switch costs were significantly negatively correlated with both the switch index ($r = -0.15$, 95% CI [−0.28, −0.01], $p = .030$) and the repetition index ($r = -0.16$, 95% CI [−0.29, −0.02], $p = .021$) from the matrix test. These correlations did not differ significantly from each other: $t(207) = -0.62$, $p = .268$.

In the structured cluster, switch costs were significantly correlated only with the interrule times of distinct-rule transitions ($r = 0.16$, 95% CI [0.01, 0.30], $p = .040$), but not with those of identical-rule transitions ($r = 0.12$, $p = .109$). In contrast, the switch costs were significantly correlated with both the switch index ($r = -0.16$, 95% CI [−0.31, −0.01], $p = .033$) and the repetition index ($r = -0.18$, 95% CI [−0.32, −0.03], $p = .020$) of the matrix test. These correlations were not significantly different from each other: $t(170) = -0.39$, $p = .348$. Mediation analyses revealed that interrule times mediated the relationship between switch costs of the task-switching paradigm and matrix performance within the structured cluster only for distinct-rule transitions: the indirect effect was significant, $\beta = -0.07$, 95% CI [−0.15, −0.01], $p = .040$ (Fig. 4). No significant mediation effects were found for identical-rule transitions in the structured cluster, nor for either condition in the unstructured cluster (all $p > .050$). However, also in the unstructured cluster, the correlations between switch costs on the one hand and the switch index ($r = -0.28$, 95% CI [−0.59, −0.06], $p = .034$) as well as the repetition index ($r = -0.29$, 95% CI [−0.61, −0.07], $p = .029$) on the other hand were significant. Again, these correlations did not differ significantly: $t(35) = -0.08$, $p = .468$.

*2.2.4. Construct validity between the two matrix indices*

The non-latent correlation between the switch index and the repetition index was strong, $r = 0.97$, 95% CI [0.96, 0.98], $p < .001$. In the CFAs, both the single-factor and the two-factor models showed acceptable overall fits (see Hu & Bentler, 1999; Kline, 2023; MacCallum, Browne, & Sugawara, 1996), with a slight advantage for the single-factor solution (Table 1). Factor loadings were consistently high in both models (all $\lambda \geq 0.74$; Fig. 5). In the two-factor model, the latent factors for the performance on distinct-rule and identical-rule transitions were almost perfectly correlated: $r = 0.99$, $p < .001$. The model comparison revealed no significant advantage of retaining two separate factors over a parsimonious single-factor solution: $\Delta\chi^2 = 0.49$, $p = .483$.

*2.2.5. Do both indices predict educational criteria similarly?*

Both the switch index ($r = -0.16$, 95% CI [−0.30, −0.03], $p = .021$) and the repetition index ($r = -0.19$, 95% CI [−0.32, −0.05], $p = .010$) from the matrix test were significantly negatively correlated with GPA (note that in the German grading system, lower values indicate better academic performance). The two correlations did not differ significantly: $t(192) = -1.29$, $p = .197$. Accordingly, the switch index did not explain additional variance in GPA beyond the repetition index ($\Delta R^2 = 0.49\%$, $F(1, 192) = 0.98$, $p = .324$). Similarly, both indices were

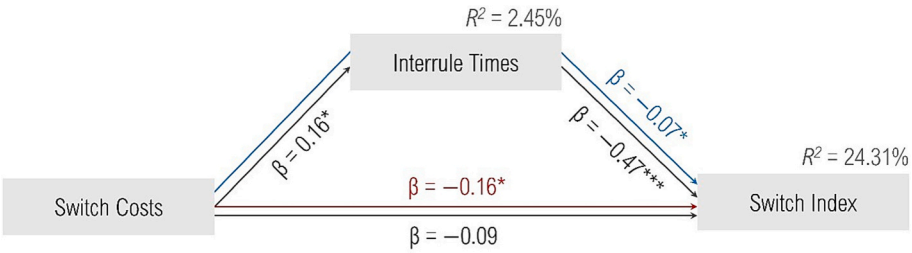



**Fig. 4.** Results of the mediation analysis within the structured cluster. Red path = total effect, black paths = direct effects, blue path = indirect (mediated) effect; the relationship between task-switching ability (switch costs) and matrix performance on distinct-rule transitions (switch index) is mediated by interrule times between matrix rules. Together, task-switching ability and interrule times accounted for 24.31% of the variance in matrix performance; * $p < .050$, *** $p < .001$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 1**
Fit Indices of the Single- and Two-Factor Model.

| Model | $\chi^2$ | *df* | $\chi^2/df$ | *p* | CFI | TLI | RMSEA | SRMR |
|---|---|---|---|---|---|---|---|---|
| Threshold | – | – | ≤ 3.00 | ≥ 0.050 | ≥ 0.95 | ≥ 0.95 | ≤ 0.10 | ≤ 0.08 |
| 1F | 23.02 | 9 | 2.56 | 0.006 | 0.98 | 0.97 | 0.09 | 0.02 |
| 2F | 22.53 | 8 | 2.82 | 0.004 | 0.98 | 0.97 | 0.09 | 0.02 |

*Note*. 1F = single-factor model, 2F = two-factor model, *df* = degrees of freedom, *p* = *p*-value, CFI = Comparative Fit Index, TLI = Tucker-Lewis Index, RMSEA = Root Mean Square Error of Approximation, SRMR = Standardized Root Mean Square Residual.

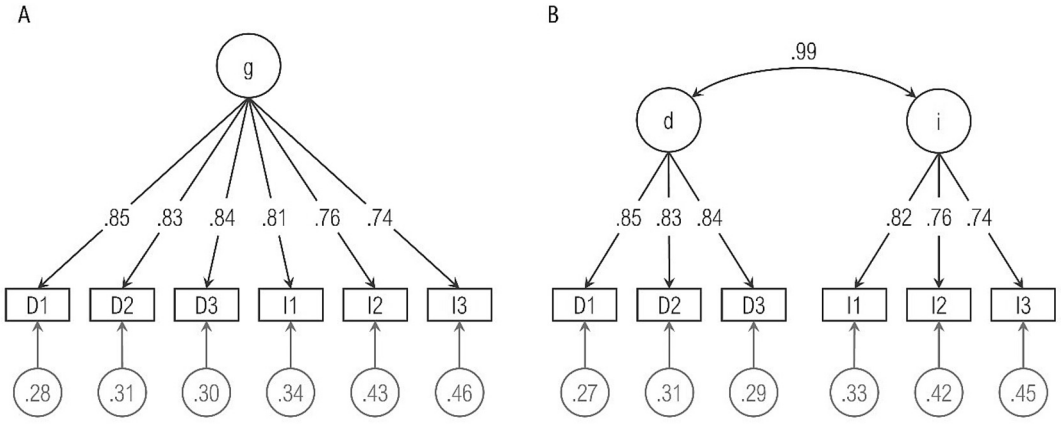



**Fig. 5.** Results of the confirmatory factor analyses. A. single-factor model: item parcels of both the distinct-rule transitions (D1–3) and identical-rule transitions (I1–3) show high loadings on a common factor (g). B. two-factor model: item parcels of the distinct-rule transitions and identical-rule transitions show high factor loadings on two separate factors (d, i) with a near-perfect latent correlation; factor loadings are standardized.

significantly correlated with the level of education: $r = 0.15$, 95% CI [0.01, 0.28], $p = .032$ for the switch index, and $r = 0.14$, 95% CI [0.01, 0.28], $p = .037$ for the repetition index. Again, the correlations did not differ significantly ($t(206) = 0.22$, $p = .826$), and the switch index did not explain incremental variance beyond the repetition index ($\Delta R^2 = 0.11\%$, $F(1, 206) = 0.63$, $p = .635$).

### 2.2.6. Post-hoc and robustness analyses

To contextualize the results, we conducted post-hoc sensitivity analyses to determine the minimal detectable effect sizes at 80% statistical power. In the full sample ($n = 209$), correlations between task-switching ability and matrix processing time or matrix performance of $r = 0.17$ were detectable with 80% power. In the structured cluster ($n = 172$), correlations of $r = 0.19$ were detectable, whereas substantially larger correlations of $r = 0.40$ would have been required in the unstructured cluster ($n = 37$) to reach 80% power.

To examine the robustness of the cluster solution, we conducted additional analyses using alternative clustering specifications. The resulting cluster assignments were highly consistent with the original solution obtained using Ward's method with Euclidean distance (Adjusted Rand Index = 1.00 for Mahalanobis distance, 0.84 for complete linkage, and 0.93 for k-means), corresponding to 96.65–100% identical classification.

In preparation for the robustness analyses based on the full sample, we compared the characteristics of included and excluded participants. The proportion of female and male participants did not differ significantly between the two groups, $\chi^2(1) = 2.13$, $p = .144$, and the groups did not differ in age, $t(87.57) = -0.32$, $p = .751$. However, included participants showed a higher GPA, $t(38.87) = -2.24$, $p = .027$, and substantially higher matrix test performance, $t(257.41) = 23.73$, $p < .001$ (for a detailed comparison, please refer to Appendix B). To examine whether the 34 participants who reported guessing in the matrix test did so because they ran out of time, we analyzed their response time patterns in more detail. The mean ToT per matrix was $M = 21.84$ s ($SD = 15.95$) despite a maximum response time of 75 s. On average the limit was reached for only $M = 1.12$ ($SD = 3.33$) out of 20 items before response was completed.

Using the full sample, two clusters with a pattern very similar to that observed in the main analyses could be extracted. As in the main analyses, the structured cluster ($n = 175$) invested more time between rules ($t(250.41) = 14.17$, $p < .001$, $d = 1.69$, 95% CI [1.42, 1.96]) and exhibited fewer rule jumps ($t(167.03) = -9.50$, $p < .001$, $d = -1.25$, 95% CI [−1.51, −1.00]) than the unstructured cluster ($n = 118$; please note that $n = 9$ participants could not be included in the clustering because no values for interrule times were available). Within the structured cluster, interrule times accounted for an additional $\Delta R^2 = 44.52\%$ of the variance in matrix performance beyond the ToT ($F(1, 172) = 155.30$, $p < .001$). Furthermore, participants spent more time between distinct than identical matrix rules ($\Delta M = 0.90$ s, $t(283) = 3.63$, $p < .001$, $d = 0.22$, 95% CI [0.10, 0.33]). In contrast to the main analyses, switch costs from the task-switching paradigm were no longer correlated with matrix performance in the full sample (switch index: $r = -0.06$, $p = .300$; repetition index: $r = -0.07$, $p = .243$). However, the assignment of participants to the structured cluster converged perfectly with the main analyses (100% identical classification), with the consequence that the functional role of task switching during structured processing of distinct matrix rules remained unchanged. For the unstructured cluster, convergence between the main and robustness analyses was 92.50%. Notably, 99.66% of the participants excluded from the main analyses were assigned to the unstructured cluster in the robustness analyses. Consistent with the main analyses, no functional role of task-switching ability was observed for unstructured matrix processing.

Replicating the results of the main analyses, both the switch index ($r = -0.20$, 95% CI [−0.31, −0.08], $p = .001$) and the repetition index ($r = -0.20$, 95% CI [−0.32, −0.08], $p = .001$) were correlated with GPA in the unfiltered sample. The same pattern emerged for the association with level of education ($r = 0.18$, 95% CI [0.06, 0.29], $p = .002$; $r = 0.18$, 95% CI [0.06, 0.28], $p = .001$).

### 2.3. Discussion

Previous research has shown moderate, yet consistent correlations between task-switching ability and matrix-test performance (e.g., Li et al., 2019; Salthouse et al., 1998; Yehene & Meiran, 2007). Study 1 aimed to clarify whether this association reflects a causal relationship or merely a correlational overlap. Given that traditional figural matrices typically require the solution of multiple distinct rules, it is plausible that mental task-switching is involved when transitioning from one rule to the next. We therefore modified the standard matrix format to include both distinct and identical rules within a single item. This experimental manipulation allowed us to examine whether task-switching ability is functionally involved in matrix processing or not.

Because a potential causal link may only emerge in individuals who process matrices in a structured and planful manner, we applied and successfully replicated the cluster analysis approach by Weber et al. (2025), classifying participants into two clusters with structured or unstructured matrix processing. From this, three key findings emerged: (1) Task-switching ability was significantly associated with matrix performance, independent of whether participants processed matrices structurally or whether they processed distinct or identical rules. (2) Task-switching ability was related to interrule times (i.e., latency between processing two consecutive matrix rules), but only for structured solvers and only on traditional distinct-rule transitions, not on identical-rule transitions. (3) In case of structured matrix processing and distinct rules, the relationship between task-switching ability and matrix performance was mediated by interrule times. In case of identical rules, no mediation was observed.

These findings suggest that both the shared-resource hypothesis and the switch-dependency hypothesis hold true: The shared-resource hypothesis is supported by the observation that the correlation between task-switching and matrix performance persisted even in the absence of actual distinct-rule transitions, indicating an underlying common cognitive mechanism (e.g., attention control). However, to substantiate this hypothesis more directly, future research could assess the potential shared resource and examine whether the correlation between task-switching ability and matrix-test performance diminishes when it is statistically controlled. In contrast, the switch-dependency hypothesis receives support from the finding that task-switching ability has a direct impact on interrule times and an indirect one on matrix performance, but only when distinct rules are processed in a structured manner (but see section 4.2 for limitations regarding statistical power in the unstructured cluster).

However, even in case of a this functional link, the correlation between task-switching ability and matrix performance did not increase compared to the non-functional condition. This suggests that the traditional matrix format (based solely on distinct logical rules) does not amplify the correlation but instead engages task-switching ability through a different cognitive pathway. While this finding is theoretically important for understanding the interplay between task-switching and matrix processing, it also carries implications for the design of figural matrices in diagnostic settings. In our study, performance on identical rules did not differ significantly from performance on distinct rules. Moreover, both performance indices correlated with GPA and level of education similarly, with no significant differences in predictive validity. This raises the issue of whether rule diversity in figural matrices holds diagnostic relevance, or whether identical rules within a single item may be used without compromising validity.

## 3. Study 2

The results of study 1 revealed a high correlation between participants' performance on distinct and that on identical matrix rules. To rule out the possibility that this correlation was simply due to both performance indices being derived from the same items, study 2 aimed to assess the convergent validity between a traditional matrix format (i.e., with distinct rules only) and the new matrix design developed in study 1 (i.e., including both distinct and identical rules). Building on these results, we sought to address RQ 2 of whether it is diagnostically justifiable to allow for identical rules within a single matrix item. To investigate this, we tested five hypotheses (H1–H5), which we preregistered on OSF prior to data collection: https://osf.io/7q5vf/?view_only=4efeab432bd14485bb64dbf53c06841e.

H1. We expected to replicate the high correlation between the two performance indices (switch index and repetition index) found in study 1.

H2. We expected both matrix tests (traditional and newly designed) to show similar test and item characteristics. Specifically, we assumed high and comparable internal consistencies, as well as similar means, standard deviations, item difficulties, and discrimination parameters.

H3. We expected a strong convergent correlation between the overall performances on the traditional and the newly designed matrix test.

H4. We expected that both tests would load on a common latent factor and that this model would provide at least an equivalent fit compared to a two-factor model with separate loadings.

H5. Prior research has revealed substantial correlations ($r$ = 0.13–0.32) of construction-based matrix tests with educational criteria (Becker & Spinath, 2014; Krieger et al., 2022; Weber et al., 2023). In study 1, we replicated this correlation. Hence, we expected both tests of this study to correlate similarly with external educational criteria.

To relate the findings of the present study to those of study 1, we conducted post-hoc analyses examining whether the association between the two test formats varied as a function of participants' processing strategy. These analyses were particularly relevant for evaluating whether the two formats can be used interchangeably across samples that differ in the prevalence of structured versus unstructured processing strategies, which may be especially important for assessments involving high-performing test-takers (e.g., student selection tests).

### 3.1. Methods

#### 3.1.1. Sample and exclusion criteria

We recruited $N = 307$ participants between June and August 2025 who participated in all parts of the study. As in study 1, participants were acquired at the local university and on SurveyCircle. As preregistered, we excluded 39 participants who reported in a post-study questionnaire that they had guessed while solving the figural matrices. Additionally, 10 participants were excluded due to implausibly short processing times on the matrix items. This resulted in a final sample of $N = 258$ participants. Participants had a mean age of $M = 27.68$ years ($SD = 7.91$ years), 62.40% identified as female, 33.72% as male, 0.78% as non-binary, and 3.10% did not state any gender. The sample showed a heterogeneous educational background (Appendix A). Participants from the local university received course credit for participation. We conducted both studies reported in this article under the same project title and participants were required to provide an individual (but anonymized) participation code to ensure that participants did not take part in both studies. This procedure allowed us to verify that the samples were independent.

#### 3.1.2. Test procedure and materials

Participants completed two figural matrix tests, each comprising 12 items. One test was an item selection of the newly designed test from study 1, which included both identical and distinct rules within a single matrix (*mixed-rules test*). Based on item difficulty and item-rest correlations, we selected three items with three rules, six with four rules, and

three with five rules. The second test followed the traditional item format, in which all rules within a matrix were distinct (*distinct-rules test*). To ensure comparability between the two tests, we constructed item pairs that shared two distinct rules alongside further different rules, and we approximately balanced the overall distribution of rule types across both tests. We used a within-subject design with randomized test order. Participants had 75 s to solve each item. After completing the cognitive tasks, participants were asked to report demographic information.

#### 3.1.3. Statistical analysis

To address H1, we computed two performance indices from the mixed-rules test analogous to the procedure used in study 1: one for distinct-rule transitions (switch index) and one for identical-rule transitions (repetition index). We then calculated the correlation between the two indices. To test H2, we compared the internal consistencies of the two tests by calculating Cronbach's alpha and examining the overlap of their 95% confidence intervals. We then compared the mean scores using a dependent *t*-test. In addition, we fitted a common polytomous graded-response model item-response theory (GRM IRT) model based on the partial solutions of the items in both tests and extracted item difficulties and discrimination parameters. We then tested the mean difference of each parameter type against zero. Moreover, we computed the Test Item Functions (TIFs) and compared their characteristics across the two tests.

To evaluate H3 regarding convergent validity, we computed the correlation between the overall performance (partial solution score) on the distinct-rules test and the mixed-rules test. We also calculated a partial correlation to control for potential order effects (i.e., distinct-rules test first vs. mixed-rules test first) and then applied a correction for attenuation to obtain the highest possible estimate of the true-score correlation (Spearman, 1904). Furthermore, we estimated person ability parameters from two separate polytomous GRM IRT models based on each test and calculated their correlation. To assess whether performance on specific rule types (e.g., addition, intersection) generalized across tests, we calculated rule-type-specific scores for each participant and correlated those between tests. To ensure that any convergence was not merely driven by the shared rules of the item pairs, we computed separate scores for shared versus unshared rules and compared their respective correlations.

To test H4, we conducted confirmatory factor analyses comparing a two-factor model (with one factor per test) to a hierarchical single-factor model, in which both test-specific factors loaded on a higher-order general factor. Therefore, we created three item parcels per test, each containing four items. We compared these two models in a model comparison test. Although item parceling has been described as a reasonable approach for unidimensional scales (e.g., Bandalos & Finney, 2001; Little et al., 2002), we conducted additional post-hoc analyses at the item level to ensure robustness of the proposed factor models.

To address H5, we computed the correlations between each matrix test and GPA and compared them using a t-test for dependent correlations.

In addition, we conducted post-hoc moderation analyses to examine whether the association between the two test formats varied as a function of processing strategy. Participants were classified into structured versus unstructured clusters using the same procedure as in study 1. Furthermore, as in study 1, we conducted robustness analyses using the full sample without case exclusions in order to contextualize the results of the main analyses.

All analyses were conducted in *R* (R Core Team, 2025) using the packages *dplyr* (Wickham et al., 2023) and *stringr* (Wickham, 2023) for data preparation, the packages *lavaan* (Rosseel, 2012), *mirt* (Chalmers, 2012), and *psych* (Revelle, 2025) for statistical analysis, and the packages *ggplot2* (Wickham, 2016), *patchwork* (Pedersen, 2025), and *purrr* (Wickham & Henry, 2025) for data visualization.

**Table 2**
Descriptive statistics of the two tests.

| Item | Number of Rules | Distinct-Rules Test | Mixed-Rules Test |
|---|---|---|---|
| 1 | 3 | 2.20 (1.09) | 2.30 (1.08) |
| 2 | 3 | 2.28 (1.01) | 2.45 (0.97) |
| 3 | 3 | 2.12 (1.07) | 2.05 (1.03) |
| 4 | 4 | 2.52 (1.33) | 2.64 (1.42) |
| 5 | 4 | 2.80 (1.45) | 2.90 (1.49) |
| 6 | 4 | 2.39 (1.41) | 2.23 (1.41) |
| 7 | 4 | 2.24 (1.42) | 2.89 (1.42) |
| 8 | 4 | 2.60 (1.44) | 2.60 (1.51) |
| 9 | 4 | 2.78 (1.44) | 2.20 (1.54) |
| 10 | 5 | 2.98 (1.78) | 3.03 (1.79) |
| 11 | 5 | 2.91 (1.80) | 2.46 (1.73) |
| 12 | 5 | 3.13 (1.80) | 2.80 (1.92) |
| Total | 48 | 30.96 (14.61) | 30.55 (14.42) |

*Note*. Mean outside brackets, standard deviation inside brackets.

### 3.2. Results

#### 3.2.1. Convergence of performance indices in the mixed-rules test

Participants achieved a switch index of $M = 0.37$ ($SD = 0.23$) and a repetition index of $M = 37$ ($SD = 0.23$). We replicated the high correlation of the two indices found in study 1, $r = 0.97$, 95% CI [0.96, 0.98], $p < .001$, indicating a strong convergence.

#### 3.2.2. Comparison of test and item characteristics

Table 2 presents the descriptive statistics of the distinct-rules and the mixed-rules test. The internal consistencies of both tests were high with overlapping confidence intervals: $\alpha_{distinct} = 0.96$, 95% CI [0.96, 0.97]; $\alpha_{mixed} = 0.95$, 95% CI [0.95, 0.96]. Neither the mean test performances, $\Delta M = 0.41$, $t(257) = 0.91$, $p = .361$, nor the standard deviations ($SD_{distinct} = 14.61$, $SD_{mixed} = 14.42$), differed significantly between the two tests.

Fig. 6 illustrates the Item Characteristic Curves (ICCs) from the polytomous GRM IRT model, plotted for each item pair (e.g., Item 1 of the distinct-rules test vs. Item 1 of the mixed-rules test). Across all 12 item pairs, the ICCs showed highly similar locations on the ability scale and comparable slopes or distributions, respectively.

Mean differences in item difficulty parameters did not significantly deviate from zero, $t(11) = -0.01$, $p = .926$, nor did differences in discrimination parameters, $t(11) = 0.17$, $p = .142$. Fig. 7 displays the TIFs of both tests, which showed highly similar locations and shapes. Both tests provided excellent psychometric quality with maximum information at similar points on the latent ability scale (distinct-rules test: $\theta = -0.77$ | mixed-rules test: $\theta = -0.61$), with comparable peaks ($I_{max} = 25.86 \mid 22.12$), similar minimum standard errors ($SEM_{min} = 0.20 \mid 0.21$), as well as near-identical and broad ranges with a local reliability of $\rho \geq 0.80$ ($I > 4$: [−2.48, 1.72], [−2.48, 1.77]).

#### 3.2.3. Convergent validity of the tests

The correlation between the overall scores on the two matrix tests was very high: $r = 0.87$, 95% CI [0.84, 0.90], $p < .001$. After controlling for order effects and correcting for attenuation, the partial correlation was $r = 0.93$. Likewise, the person ability estimates derived from the respective GRM IRT models were strongly correlated: $r = 0.85$, 95% CI [0.82, 0.88], $p < .001$. Fig. 8 shows scatterplots illustrating (A) the correlation between the overall test scores and (B) the correlation between the estimated IRT ability parameters. Additionally, rule-type-specific scores of both tests correlated strongly: $r_{addition} = 0.78$ (95% CI [0.73, 0.82]), $r_{subtraction} = 0.80$ (95% CI [0.75, 0.84]), $r_{rotation} = 0.75$ (95% CI [0.69, 0.80]), $r_{intersection} = 0.80$ (95% CI [0.75, 0.84]), $r_{extinction} = 0.80$ (95% CI [0.75, 0.84]), all $p < .001$. To rule out that convergence was driven solely by the shared rules of the item pairs, we separately computed correlations for rule scores based on shared and unshared rules. These correlations were similarly high (Table 3), indicating that the observed convergence extended beyond the overlap in some rules.

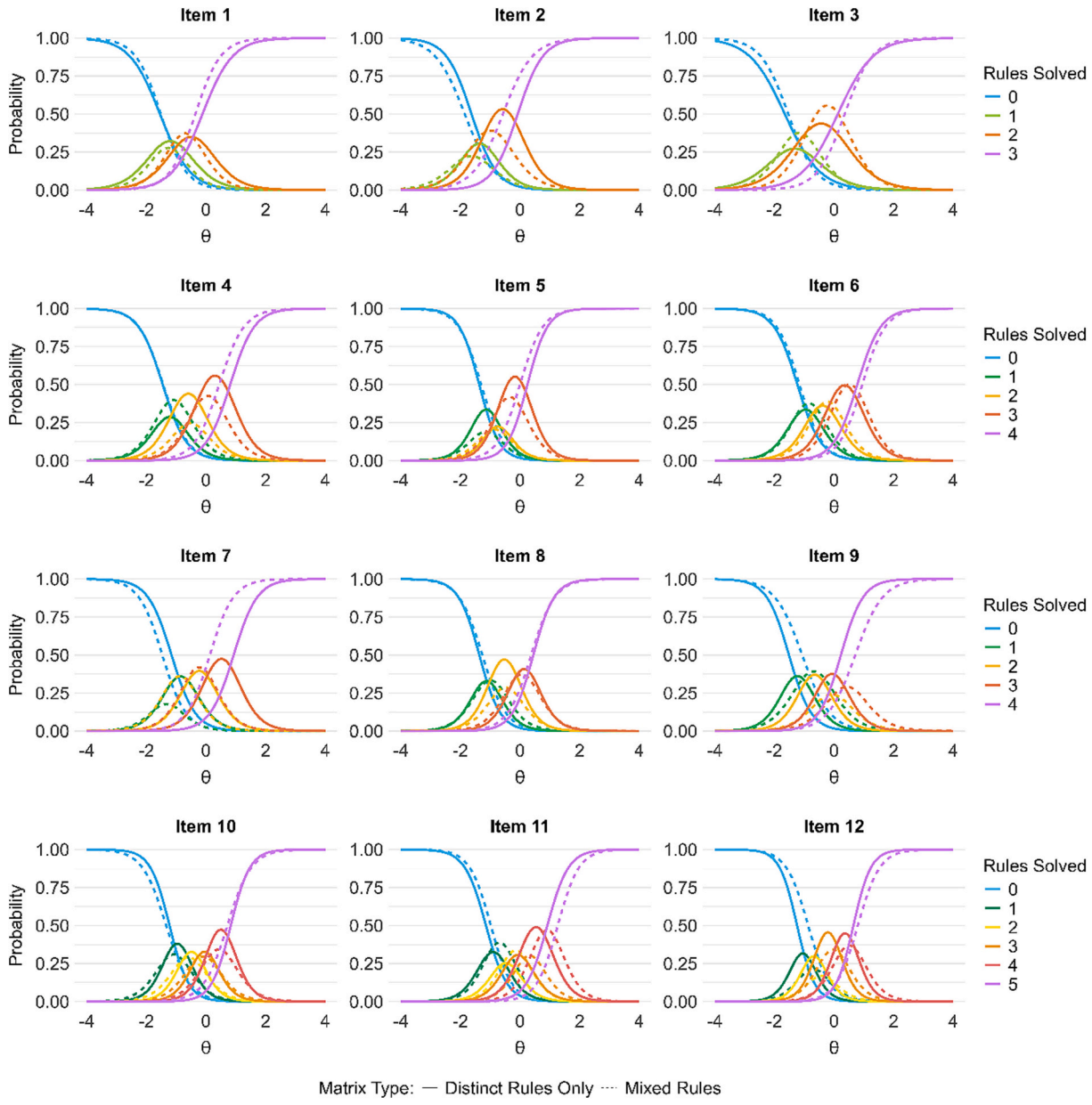



**Fig. 6.** Item characteristic curves (ICCs) for the distinct-rules test (solid lines) and the mixed-rules test (dashed lines), based on a polytomous GRM IRT model. The curves represent the probability of solving a given number of rules as a function of participant's latent ability. Across the twelve items, no systematic deviations in item difficulty or discrimination were observed between the two tests, indicating comparable psychometric properties.

### 3.2.4. *Latent model comparison*

Confirmatory factor analyses revealed acceptable model fit for both the hierarchic single-factor model and the two-factor model (Table 4). The models did not differ significantly in fit, $\Delta\chi^2 = 0.00$, $p = 1.000$, indicating that a single general factor underlying both matrix formats is sufficient to account for the observed covariance structure.

### 3.2.5. *External validity of the tests*

A total of 225 participants reported an academic high school GPA. On average, they had a GPA of $M = 2.06$ with a standard deviation of $SD = 0.63$ (note that the German grading system regarding successful GPA ranges from 1 to 4 where lower grades indicate better school achievement). The correlation between the mixed-rules test and GPA was significant and consistent with the findings from study 1: $r = -0.23$, 95% CI [−0.35, −0.10], $p = .001$. Also, the traditional distinct-rules test was significantly correlated with GPA: $r = -0.23$, 95% CI [−0.35, −0.10], $p = .001$. The strength of the two correlations did not differ significantly, $t(223) = 0.06$, $p = .950$, suggesting that both formats have comparable predictive validity for academic achievement.

### 3.2.6. *Post-hoc and robustness analyses*

The post-hoc clustering analysis replicated the two-cluster pattern observed in study 1, consisting of a structured ($n = 204$) and an unstructured cluster ($n = 54$). The correlation between the two test formats was very similar within the structured ($r = 0.88$, 95% CI [0.85, 0.91], $p < .001$) and within the unstructured cluster ($r = 0.83$, 95% CI [0.73, 0.90], $p < .001$), and the cluster membership did not significantly moderate the association between the two formats ($\beta = 0.01$, 95% CI [−0.14, 0.17], $p = .887$).

Robustness analyses of the factorial structure confirmed the results of the main analyses at the item level. The hierarchical model, which subsumed the two matrix formats under a global factor, again showed an acceptable to good fit: $\chi^2(250) = 582.98$, $p < .001$, $\chi^2/df = 2.32$, CFI = 0.95, TLI = 0.94, RMSEA = 0.07, SRMR = 0.04, all factor loadings $\lambda >$

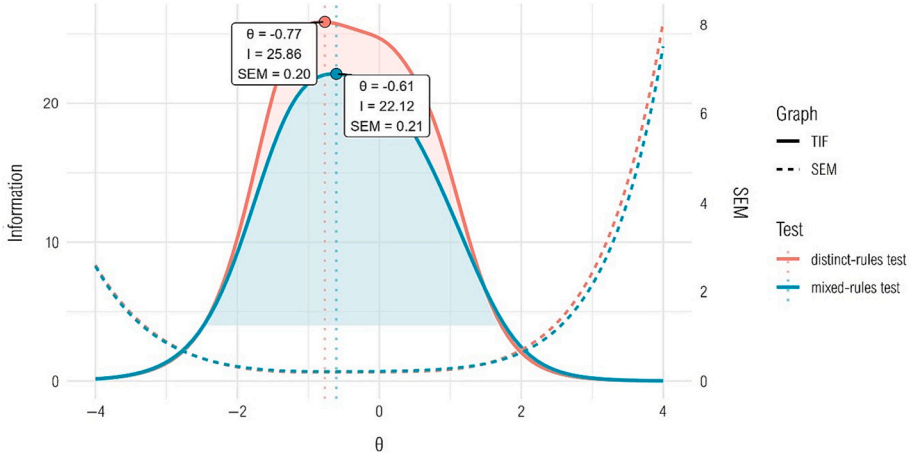



**Fig. 7.** Test information functions (TIFs) and standard error (SEM) curves of the two tests based on polytomous GRM IRT models. TIFs are very similar regarding latent ability (θ) of maximum information, peak of information (I) and minimum SEM. Furthermore, they cover a broad and near-identical range with a local reliability of $\rho \geq 0.80$ (colored area under curve).

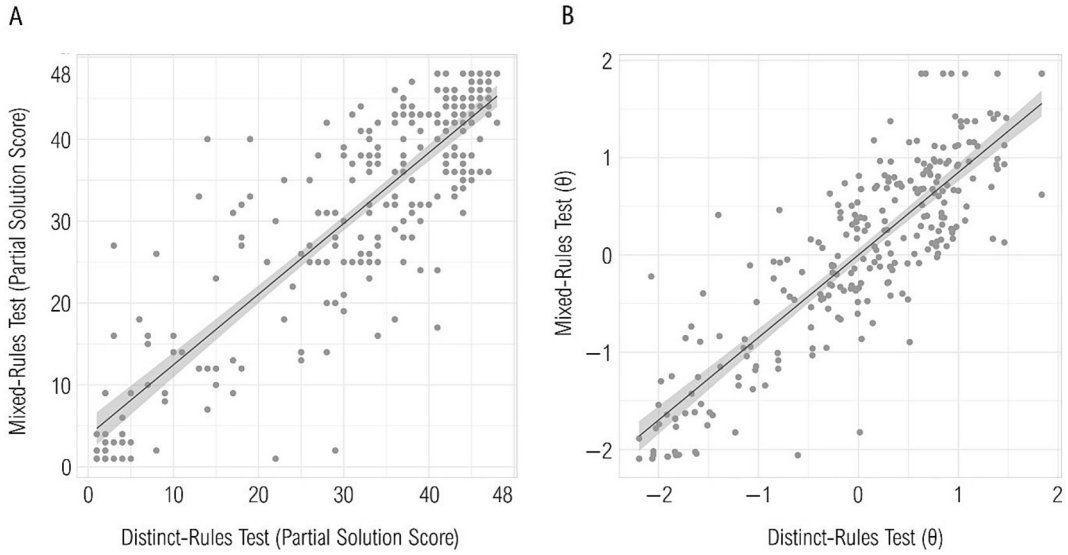



**Fig. 8.** A. Correlation between the manifest performance on the distinct-rules test and the mixed-rules test ($r = 0.87$, $p < .001$). B. Correlation between the latent ability estimates derived from the distinct-rules test and mixed-rules test based on a polytomous GRM IRT model ($r = 0.85$, $p < .001$).

**Table 3**
Correlations between shared and unshared rules.

| Distinct-Rules Test | Mixed-Rules-Test | |
|---|---|---|
| | Shared Rules | Unshared Rules |
| Shared Rules | 0.84*** [0.80, 0.87] | 0.84*** [0.81, 0.88] |
| Unshared Rules | 0.85*** [0.81, 0.88] | 0.87*** [0.84, 0.90] |

*Note.* *** $p < .001$. Values in square brackets indicate the 95% confidence interval. The correlation did not depend on whether rules were shared across tests.

0.70. In the two-factor model with separate factors for each matrix test, the two factors correlated strongly ($r = 0.91$, $p < .001$), and the two-factor solution did not provide a better fit than the hierarchical single-factor model ($\Delta\chi^2 = 0.00$, $p = 1.000$).

In preparation for the robustness analyses using the unfiltered sample, we compared the characteristics of included and excluded participants (Appendix B). Female and male participants were equally distributed across the two groups, $\chi^2(1) < 0.001$, $p = 1.000$, and the groups did not differ in age, $t(50.17) = 0.48$, $p = .636$. Likewise, no difference was observed in GPA ($t(38.87) = 1.20$, $p = .237$). However, included participants achieved substantially higher matrix performance

**Table 4**
Fit indices of the hierarchic single- and two-factor model.

| Model | $\chi^2$ | *df* | $\chi^2/df$ | *p* | CFI | TLI | RMSEA | SRMR |
|---|---|---|---|---|---|---|---|---|
| Threshold | – | – | $\leq 3.00$ | $\geq 0.050$ | $\geq 0.95$ | $\geq 0.95$ | $\leq 0.10$ | $\leq 0.08$ |
| h-1F | 22.21 | 7 | 3.17 | 0.002 | 0.99 | 0.99 | 0.09 | 0.01 |
| 2F | 22.21 | 8 | 2.78 | 0.005 | 0.99 | 0.99 | 0.08 | 0.01 |

*Note.* h-1F = hierarchic single-factor model, 2F = two-factor model, *df* = degrees of freedom, *p* = *p*-value, CFI = Comparative Fit Index, TLI = Tucker-Lewis Index, RMSEA = Root Mean Square Error of Approximation, SRMR = Standardized Root Mean Square Residual.

than excluded participants, $t(115.25) = 1.20$, $p < .001$. Post-hoc analyses of the response times of participants who reported guessing indicated that they spent on average only $M = 21.48$ s ($SD = 17.76$) per item despite the maximum response time of 75 s, and that timeouts occurred for an average of $M = 5.01$ ($SD = 5.39$) out of 24 items. This pattern suggests that guessing was unlikely to have occurred merely because participants ran out of time.

Robustness analyses using the full unfiltered sample ($n = 309$) again revealed a very strong association between the distinct-rules and the mixed-rules test: $r = 0.91$, 95% CI [0.89, 0.93], $p < .001$. Performance in both the distinct-rules test ($r = -0.22$, 95% CI [−0.33, −0.10], $p < .001$) and the mixed-rules test ($r = -0.22$, 95% CI [−0.33, −0.10], $p < .001$) remained substantially and equally correlated with GPA.

### 3.3. Discussion

Study 1 demonstrated that task-switching ability can play a necessary role in solving figural matrices when transitioning between distinct logical rules as in traditional matrix tests. However, this necessary involvement did not increase the strength of the correlation between task-switching ability and matrix-test performance compared to conditions involving identical (i.e., repeated) rules. Building on these theoretical findings, study 2 addressed the practical implications for test development and diagnostic application. Specifically, we examined whether allowing for identical logical rules in a single item compromises the psychometric quality of a matrix test. To this end, participants completed two tests: one based on the traditional matrix design (with distinct rules only) and one based on the re-designed format from study 1 (including both distinct and identical rules per item). Across all analyses, we observed strong evidence for convergent validity between the two formats. The correlation between the overall performances was $r = 0.87$, and the correlation between the IRT-based ability estimates was $r = 0.85$. Furthermore, item-level characteristics, test information, and factorial validity were highly comparable across both formats. Finally, the correlations with GPA were in line with prior findings on the external validity of construction-based figural matrices (Becker & Spinath, 2014; Krieger et al., 2022; Weber et al., 2023). Taken together, these findings suggest that, from a diagnostic perspective, it is not necessary to restrict matrix items to distinct rules only. The inclusion of identical rules within an item does not appear to compromise reliability, construct validity, or external validity, offering greater flexibility for item construction without diagnostic costs.

## 4. General discussion

### 4.1. Consolidation and interpretation of the results

The overarching goal of this research was to examine whether the use of distinct logical rules as in traditional figural matrix tests plays cognitively and diagnostically a necessary role. To this end, we investigated (a) the nature of the link between task-switching ability and matrix processing, and (b) the diagnostic consequences of relaxing the distinct-rule constraint in item construction. Across two complementary studies, we tested both the theoretical rationale for rule switching and its practical implications for test development.

In study 1, we investigated the nature of the link between task-switching ability and matrix-test performance. Previous research has consistently demonstrated a moderate association between task-switching ability and matrix-test performance (e.g., Li et al., 2019; Salthouse et al., 1998; Yehene & Meiran, 2007). The present study advances this literature by disentangling the functional structure underlying this association. By employing a novel item format that varies rule diversity within matrices and systematically analyzing log file data, we were able to link process data with established theoretical frameworks (cf. Stadler, Brandl, & Greiff, 2023), thereby isolating the functional role of task-switching ability in figural matrix processing. Our findings suggest that task-switching ability exerts a necessary influence when test-takers process distinct matrix rules. Analogous to task-set inertia in task-switching paradigms (e.g., Kiesel et al., 2010; Schmitz & Krämer, 2023), traditional matrix items appear to induce a kind of *rule inertia*: transitioning from one logical rule to another requires more time when rules are distinct and hence may reflect proactive interference between rules. This supports the view that distinct-rule transitions challenge cognitive flexibility and disengagement processes. However, even when such rule switching is not required (i.e., when identical rules are processed consecutively) the correlation between task-switching ability and matrix performance remains similarly strong. This dual pattern supports both the shared-resource hypothesis (that task-switching and matrix processing draw on a common underlying cognitive resource) and the switch-dependency hypothesis (which posits a functional role of task-switching in distinct-rule transitions). When necessary for item processing, task-switching ability appears to take another pathway.

In study 2, we examined whether allowing for identical rules within a single matrix item compromises psychometric validity. Across multiple indicators (test score correlations, IRT parameters, factorial structure, and external validity) the results showed very high convergent validity between the traditional matrix format (with only distinct rules) and the newly developed mixed-rule format. The association between the two matrix formats remained stable across both processing strategies (structured versus unstructured), indicating that the formats can be used equivalently regardless of the specific composition of the sample.

### 4.2. Limitations

Both samples in the present article were relatively heterogeneous with respect to educational backgrounds (see Appendix A). In study 1, participants represented all eleven fields of education defined by UNESCO, whereas the participants of study 2 represented nine of these fields. However, it should be noted that a substantial number of participants were excluded: $n = 93$ of the initial 302 participants in study 1 and $n = 49$ of the initial 307 participants in study 2. These exclusions can be attributed in part to the deliberately liberal initial inclusion criteria, which required only that participants provided at least one response in each part of the assessment. Based on preregistered criteria, $n = 34$ and $n = 39$ participants were excluded because they reported in a questionnaire that they had not engaged seriously in solving the matrices. Additional $n = 46$ and $n = 10$ participants were excluded due to implausibly fast response times which is likely attributable to the comparatively long duration of the assessment, particularly in study 1 (approximately 50 min). Post-hoc analyses based on the full sample without case exclusions suggested that the core findings of both studies were robust.

A second limitation concerns the association between task-switching ability and matrix performance. Previous research reported correlations ranging from $r = -0.31$ to $-0.46$ (Salthouse et al., 1998; Yehene & Meiran, 2007), whereas the association observed in the present study was considerably smaller ($r = -0.15$ to $-0.16$). One possible explanation is that the sample had a relatively high overall ability level, as reflected in comparatively high level of education and matrix test scores. Consequently, range restriction may have led to an underestimation of the true association. Future studies could employ the present study design to replicate the findings in more performance-heterogeneous samples.

A third limitation concerns the imbalance in cluster size observed in study 1. We successfully replicated the cluster pattern reported by Weber et al. (2025). Their solution also exhibited unequal cluster sizes, with more participants in the structured ($n = 111$) than in the unstructured ($n = 97$) cluster. In the present study, this imbalance was substantially larger ($n = 172$ vs. 39). This pattern may be partially attributable to the sampling procedure: Part of the sample was recruited via the survey platform SurveyCircle. which allows users to earn credits by participating in other users' studies and to redeem these credits to

recruit participants for their own research. The majority of SurveyCircle users have an academic background, as the platform is primarily used for bachelor's, master's, and doctoral theses. Consequently, a large proportion of the participants recruited through SurveyCircle may have already demonstrated the ability to work in a structured manner and to meet cognitive challenges, which could have contributed to a higher prevalence of the structured cluster in the present sample. As a consequence of the small sample size in the unstructured cluster, the minimal detectable correlation with 80% statistical power was $r = \pm 0.40$. Although correlations of this magnitude would fall within the range reported in previous studies on the relationship between task-switching ability and matrix performance (e.g., Salthouse et al., 1998; Yehene & Meiran, 2007), the observed correlation in the total sample was considerably smaller. Accordingly, the present findings support a functional role of task-switching ability during structured matrix processing. However, given the limited statistical power in the unstructured cluster, a similar role during unstructured processing cannot be ruled out.

Finally, we used a GPA criterion to examine the external validity of the extracted matrix performance indices. The observed correlations of $r = -0.16$ to $-0.20$ (study 1) and $r = -0.22$ to $-0.23$ (study 2) were smaller than the meta-analytic association of (artifact-corrected) $\rho = -0.44$ between nonverbal intelligence tests and school grades (Roth et al., 2015). However, they were consistent with the magnitude reported in previous matrix-specific research ($r = -0.13$ to $-0.32$; Becker & Spinath, 2014; Krieger et al., 2022; Weber et al., 2023). Figural matrices primarily assess reasoning ability rather than the broader construct of general intelligence, which may partly explain the moderate associations with educational outcomes. In addition, school grades do not constitute purely intellectual measures but are also influenced by a range of non-cognitive factors such as personality and motivational aspects. Moreover, as discussed above, both samples were relatively high-performing in terms of academic achievement, which may have led to a range restriction and consequently to an underestimation of the association between matrix performance and educational outcomes. This assumption is supported by the finding that the correlation in study 1 increased when the previously excluded participants, who generally showed lower performance, were included in the robustness analyses. At the same time, figural matrices are frequently used in student selection tests (e.g., BaPsy; Schulz-Hardt, 2025), where similarly high-performing samples are typically assessed. In such contexts, matrix tests also show substantial convergent validity with other subtests (Levacher et al., 2023). Thus, the present findings may provide a realistic estimate of the magnitude of these associations in comparable selection settings. Furthermore, it should be noted that GPA data were obtained via self-report and may therefore be subject to recall errors or intentional misreporting. Nevertheless, prior research has shown that self-reported grades are highly correlated with official records and that absolute deviations are typically small (Sparfeldt, Buch, Rost, & Lehmann, 2008; Sticca et al., 2017).

### 4.3. Implications and future perspectives

Figural Matrices have a long-standing tradition in intelligence assessment. Their primary purpose has been to measure inductive reasoning as a core component of fluid intelligence (cf. McGrew, 2009). However, this perspective has been broadened by the view that executive functions also contribute to matrix solving: Carpenter et al. (1990) laid the groundwork for considering goal management as a key component in matrix processing (see also Embretson, 1998; Loesche et al., 2015). In line with this, it has been shown that structured (i.e., sequential rule-by-rule) processing of figural matrices is associated with higher test performance (Weber et al., 2023). Building on these insights, the present research augments the understanding of the cognitive mechanisms involved in matrix processing beyond mere rule induction: Besides goal management, also task-switching ability can play a functional role in figural matrices.

From a theoretical perspective, this research offers a novel contribution by moving beyond correlational evidence on the link between task-switching ability and matrix-test performance. By leveraging log-file based process indicators and experimentally varying rule diversity, we provide evidence that task-switching ability is not merely correlated with matrix-test performance but *can* function as a causal driver under specific processing conditions. This addresses a longstanding ambiguity in literature and supports a dual-mechanism view including both cognitive involvement and shared resource overlap. An important perspective for future research concerns the nature of the potential shared cognitive resource driving (part of) the correlation between task-switching ability and performance on figural matrix tests. Prior work suggests substantial links between WMC and both task-switching ability (e.g., Baddeley, Chincotta, & Adlam, 2001; Emerson & Miyake, 2003; Wang, Zhou, Peng, & Hu, 2022) and matrix performance (e.g., Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Domnick et al., 2017; Prabhakaran, Smith, Desmond, Glover, & Gabrieli, 1997). Shipstead et al. (2016) proposed a theoretical framework in which WMC places demands on maintenance functions and matrices on disengagement functions which are both considered a manifestation of a shared top-down executive attention. Future studies could investigate whether the observed associations between task-switching ability and matrix-test performance diminish when controlling for individual differences in executive attention.

Finally, our findings contribute to the understanding of the ToT effect in figural matrices. Previous studies have reported a quadratic relationship between ToT and matrix performance (e.g., Becker, Schmitz, Göritz, & Spinath, 2016; Goldhammer et al., 2015; Krämer et al., 2023): longer processing times are beneficial for individuals with lower cognitive ability, while shorter times go along with better performance of higher-ability individuals. Recent log-file research highlights interrule times as a critical ToT component, likely reflecting the mental construction of partial solutions. Our study extends this view by showing that interrule times are also partially influenced by task-switching ability, as evidenced by their correlation with switch costs during distinct-rule transitions. To deepen the understanding of interrule times as a central process indicator, future research may consider extending the current paradigm by incorporating (a) further cognitive measures, (b) motivational variables that influence engagement in matrix solving, and (c) personality traits such as conscientiousness.

Importantly, our findings suggest that figural matrices involving both distinct and identical logical rules engage partly different cognitive mechanisms but nevertheless hold highly comparable psychometric properties. From a test-development perspective, this implies that item construction can be made less restrictive without compromising measurement quality or diagnostic validity. This opens up practical opportunities for test development: for instance, an item pool or item generator such as the OMIB framework (Koch et al., 2022) could incorporate mixed-rule designs. This would be especially beneficial for adaptive and large-scale applications or student selection tests such as BaPsy, where a continuous supply of psychometrically sound items is essential to maintaining test validity over time.

## 5. Conclusions

This research investigated the nature of the link between task-switching ability and figural matrix performance and derived practical implications for test development. While relaxing item construction constraints by allowing identical rules within a single matrix partially alters the cognitive mechanisms involved in matrix processing, it does not compromise reliability, factorial structure, or external validity. Keeping these theoretical insights in mind, the findings broaden the degrees of freedom in item construction, which may be especially beneficial for large-scale assessments and student selection tests that require continuous item renewal.

## CRediT authorship contribution statement

**Dominik Weber:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Stella Jelen:** Writing – review & editing, Software, Methodology, Investigation, Data curation, Conceptualization. **Frank M. Spinath:** Writing – review & editing, Validation, Supervision, Resources. **Florian Krieger:** Writing – review & editing, Validation, Supervision. **Nicolas Becker:** Writing – review & editing, Validation, Supervision. **Marco Koch:** Writing – review & editing, Validation, Supervision.

## Informed consent statement

Informed consent was obtained from all subjects involved in the studies.

## Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Declaration of competing interest

The authors declare that they have no competing interests related to the publication of this research.

## Appendix A. Appendix

Table A. Educational background of the samples.

| Field of Education | Study 1 (%) | Study 2 (%) |
|---|---|---|
| Generic programmes and qualifications | 1 (0.48) | 0 (0.00) |
| Education | 9 (4.31) | 6 (2.33) |
| Arts and humanities | 8 (3.83) | 8 (3.10) |
| Social sciences, journalism and information | 55 (26.32) | 160 (62.02) |
| … of which Psychology | 8 (3.83) | 69 (26.74) |
| Business, administration and law | 8 (3.83) | 11 (4.26) |
| Natural sciences, mathematics and statistics | 17 (8.13) | 3 (1.16) |
| Information and Communication Technologies | 13 (6.22) | 16 (6.20) |
| Engineering, manufacturing and construction | 3 (1.44) | 13 (5.04) |
| Agriculture, forestry, fisheries and veterinary | 1 (0.48) | 0 (0.00) |
| Health and welfare | 7 (3.35) | 6 (2.33) |
| Services | 3 (1.44) | 6 (2.33) |
| Pupils | 47 (22.49) | 0 (0.00) |
| Not stated | 37 (17.70) | 29 (11.24) |

*Note*. According to the International Standard Classification of Education Fields of Education and Training (ISCED-F 2013) provided by UNESCO (UNESCO, 2015). As 47 participants (14.38%) in study 1 were pupils in upper secondary education (academic track) from two German federal states (Baden-Württemberg and North Rhine-Westphalia), the additional category *pupils* is reported alongside the ISCED-F categories. The remaining participants were recruited either at the local university (study 1: 22.49%; study 2: 2.93%) or via the survey platform SurveyCircle (SurveyCircle, 2025; study 1: 62.68%; study 2: 97.07%). Counts are given first, with percentage in parentheses.

Table B. Comparison of included and excluded participants.

| Variable | Included | Excluded | Statistics | *p* |
|---|---|---|---|---|
| Study 1 | | | | |
| Sex | f: 145, m: 60 | f: 50, m: 32 | $\chi^2(1) = 2.13$ | 0.144 |
| Age | 27.27 (11.02) | 27.97 (16.13) | $t(87.57) = -0.32$ | 0.751 |
| GPA | 2.08 (0.63) | 2.29 (0.65) | $t(38.87) = -2.24$ | 0.027 |
| Performance | 52.76 (19.78) | 6.80 (13.27) | $t(257.41) = 23.73$ | <0.001 |
| Study 2 | | | | |
| Sex | f: 161, m: 87 | f: 25, m: 14 | $\chi^2(1) < 0.001$ | 1.000 |
| Age | 27.68 (7.91) | 26.93 (9.56) | $t(50.17) = 0.48$ | 0.636 |
| GPA | 2.07 (0.62) | 2.25 (0.83) | $t(38.87) = -1.20$ | 0.237 |
| Performance | 61.51 (28.10) | 9.16 (15.80) | $t(115.25) = 18.33$ | < 0.001 |

*Note*. Sex is reported as absolute frequencies; all other variables are presented as means with standard deviations in parentheses. Performance refers to the total number of correctly solved matrix rules. f = female, m = male, GPA = grade point average.

Table C. Item-level statistics of the matrix test in study 1.

| Item | Switch Score | Switch Inter | Repetition Score | Repetition Inter |
|---|---|---|---|---|
| 1 | 0.95 (0.61) | 10.36 (5.56) | 0.64 (0.49) | 10.30 (6.36) |
| 2 | 0.93 (0.68) | 15.11 (8.68) | 0.48 (0.51) | 13.04 (8.45) |
| 3 | 1.09 (0.65) | 11.28 (7.45) | 0.62 (0.49) | 9.42 (6.05) |
| 4 | 0.77 (0.59) | 15.37 (8.45) | 0.61 (0.49) | 12.00 (7.88) |
| 5 | 0.55 (0.51) | 19.30 (11.30) | 0.16 (0.37) | 14.54 (10.12) |
| 6 | 0.89 (0.63) | 12.63 (5.64) | 0.44 (0.57) | 11.89 (5.28) |
| 7 | 1.09 (0.86) | 13.32 (8.15) | 0.51 (0.68) | 12.05 (7.43) |
| 8 | 0.86 (0.61) | 12.13 (7.92) | 1.29 (0.73) | 11.11 (4.74) |
| 9 | 0.94 (0.71) | 9.80 (6.03) | 1.45 (0.70) | 8.03 (3.58) |
| 10 | 1.31 (1.01) | 14.56 (7.78) | 0.58 (0.73) | 12.67 (7.09) |
| 11 | 1.14 (0.90) | 13.41 (7.03) | 0.44 (0.57) | 14.82 (8.07) |
| 12 | 1.01 (0.69) | 14.70 (7.22) | 1.36 (0.74) | 10.76 (4.93) |
| 13 | 1.03 (0.83) | 11.92 (6.95) | 1.08 (0.84) | 10.89 (5.40) |
| 14 | 0.56 (0.57) | 11.63 (8.02) | 0.57 (0.58) | 13.76 (7.44) |
| 15 | 2.16 (0.93) | 9.48 (4.49) | 0.39 (0.49) | 8.04 (3.29) |
| 16 | 1.48 (0.97) | 11.23 (5.07) | 1.13 (0.92) | 10.94 (5.39) |
| 17 | 1.08 (0.80) | 9.87 (4.67) | 1.52 (0.94) | 10.69 (5.28) |
| 18 | 1.20 (1.02) | 12.46 (7.76) | 0.89 (0.82) | 10.70 (5.73) |
| 19 | 0.91 (0.87) | 12.83 (6.47) | 1.41 (0.95) | 10.35 (5.27) |
| 20 | 1.56 (0.99) | 9.18 (4.92) | 1.51 (1.03) | 7.85 (3.09) |

*Note*. Means are given first, with standard deviations in parentheses.

## Data availability

We have shared the link to our code at the Attach File Step.

## References


Allport, A., & Wylie, G. (2000). Task switching, stimulus-response bindings, and negative priming. In S. Monsell, & J. Driver (Eds.), *Attention and performance XVIII: Control of cognitive processes XVIII* (pp. 35–70). Cambridge, MA: MIT Press.

Allport, A., Styles, E. A., & Hsieh, S. (1994). Shifting intentional set: Exploring the dynamic control of tasks. In C. Umiltà, & M. Moscovitch (Eds.), *Attention and performance XI: Conscious and nonconscious information processing* (pp. 421–452). Cambridge, MA: MIT Press.

Altmann, E. M. (2004). The preparation effect in task switching: Carryover of SOA. *Memory & Cognition, 32*, 153–163. https://doi.org/10.3758/BF03195828

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

Baddeley, A., Chincotta, D., & Adlam, A. (2001). Working memory and the control of action: Evidence from task switching. *Journal of Experimental Psychology: General, 130*(4), 641–657. https://doi.org/10.1037//0096-3445.130.4.641

Bandalos, D. L., & Finney, S. J. (2001). Item parceling issues in structural equation modeling. In G. A. Marcoulides, & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 269–296). Lawrence Erlbaum Associates, Inc.

Becker, N., & Spinath, F. M. (2014). *Design a Matrix – Advanced (DESIGMA): Ein distraktorfreier Matritzentest zur Erfassung der allgemeinen Intelligenz*. Hogrefe.

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

Becker, N., Schmitz, F., Göritz, A. S., & Spinath, F. M. (2016). Sometimes more is better, and sometimes less is better: Task complexity moderates the response time accuracy correlation. *Journal of Intelligence, 4*(3), 11. https://doi.org/10.3390/jintelligence4030011

Behrendt, S. (2023). *lm.beta: Add Standardized Regression Coefficients to Linear-Model-Objects (R Package Version 1.7-2)*. Comprehensive R Archive Network (CRAN). Available online https://CRAN.R-project.org/package=lm.beta (accessed on 10 August 2025).

Ben-Shachar, M. S., Lüdecke, D., & Makowski, D. (2020). Effectsize: Estimation of effect size indices and standardized parameters. *Journal of Open Source Software, 5*(56), 2815. https://doi.org/10.21105/joss.02815

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

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press.

Chalmers, R. P. (2012). Mirt: A multidimensional item response theory package for the r environment. *Journal of Statistical Software, 48*(6), 1–29.

Chen, G., Liu, Y., & Mao, Y. (2024). Understanding the log file data from educational and psychological computer-based testing: A scoping review protocol. *PLoS One, 19*(5), Article e0304109. https://doi.org/10.1371/journal.pone.0304109

Conway, A. R., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence, 30*(2), 163–183. https://doi.org/10.1016/S0160-2896(01)00096-4

Domnick, F., Zimmer, H. D., Becker, N., & Spinath, F. M. (2017). Is the correlation between storage capacity and matrix reasoning driven by the storage of partial solutions? A pilot study of an experimental approach. *Journal of Intelligence, 5*(2), 21. https://doi.org/10.3390/jintelligence5020021

Embretson, S. E. (1998). A cognitive design system approach to generating valid tests: Application to abstract reasoning. *Psychological Methods, 3*(3), 380–396. https://doi.org/10.1037/1082-989X.3.3.380

Emerson, M. J., & Miyake, A. (2003). The role of inner speech in task switching: A dual-task investigation. *Journal of Memory and Language, 48*(1), 148–168. https://doi.org/10.1016/S0749-596X(02)00511-9

Formann, A. K., Waldherr, K., & Piswanger, K. (2011). *Wiener Matrizen-Test 2 (WMT-2): Ein Rasch-skaldierter sprachfreier Kurztest zu Erfassung der Intelligenz*. Beltz Test: Hogrefe.

Frischkorn, G. T., & Oberauer, K. (2021). Intelligence test items varying in capacity demands cannot be used to test the causality of working memory capacity for fluid intelligence. *Psychonomic Bulletin & Review, 28*, 1423–1432. https://doi.org/10.3758/s13423-021-01909-w

Goldhammer, F., Naumann, J., & Greiff, S. (2015). More is not always better: The relation between item response and item response time in raven's matrices. *Journal of Intelligence, 3*(1), 21–40. https://doi.org/10.3390/jintelligence3010021

Harrison, T. L., Shipstead, Z., & Engle, R. W. (2015). Why is working memory capacity related to matrix reasoning tasks? *Memory & Cognition, 43*(3), 389–396. https://doi.org/10.3758/s13421-014-0473-3

Himi, S. A., Bühner, M., Schwaighofer, M., Klapetek, A., & Hilbert, S. (2019). Multitasking behavior and its related constructs: Executive functions, working memory capacity, relational integration, and divided attention. *Cognition, 189*, 275–298. https://doi.org/10.1016/j.cognition.2019.04.010

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Jensen, A. R. (1998). *The g factor: The science of mental ability*. Praeger.

Kassambara, A., & Mundt, F. (2020). *factoextra: Extract and Visualize the Results of Multivariate Data Analyses (R Package Version 1.0.7)*. Comprehensive R Archive Network (CRAN). Available online: https://CRAN.R-project.org/package=factoextra (accessed on 10 August 2025).

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching—A review. *Psychological Bulletin, 136*(5), 849–874. https://doi.org/10.1037/a0019842

Kline, R. B. (2023). *Principles and practice of structural equation modeling*. Guilford Publications.

Koch, I. (2001). Automatic and intentional activation of task sets. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(6), 1474–1486. https://doi.org/10.1037/0278-7393.27.6.1474

Koch, I. (2005). Sequential task predictability in task switching. *Psychonomic Bulletin & Review, 12*(1), 107–112. https://doi.org/10.3758/BF03196354

Koch, I., Poljac, E., Müller, H., & Kiesel, A. (2018). Cognitive structure, flexibility, and plasticity in human multitasking—An integrative review of dual-task and task-switching research. *Psychological Bulletin, 144*(6), 557–583. https://doi.org/10.1037/bul0000144

Koch, M., Spinath, F. M., Greiff, S., & Becker, N. (2022). Development and validation of the open matrices item bank. *Journal of Intelligence, 10*(3), 41. https://doi.org/10.3390/jintelligence10030041

Krämer, R. J., Koch, M., Levacher, J., & Schmitz, F. (2023). Testing replicability and generalizability of the time on task effect. *Journal of Intelligence, 11*(5), 82. https://doi.org/10.3390/jintelligence11050082

Krieger, F., Zimmer, H. D., Greiff, S., Spinath, F. M., & Becker, N. (2019). Why are difficult figural matrices hard to solve? The role of selective encoding and working memory capacity. *Intelligence, 72*, 35–48. https://doi.org/10.1016/j.intell.2018.11.007

Krieger, F., Becker, N., Greiff, S., & Spinath, F. M. (2022). *Design a Matrix – Standard (DESIGMA): Ein distraktorfreier Matritzentest zur Erfassung der allgemeinen Intelligenz.* Hogrefe.

Levacher, J., Koch, M., Stegt, S. J., Hissbach, J., Spinath, F. M., Escher, M., & Becker, N. (2023). The construct validity of the main student selection tests for medical studies in Germany. *Frontiers in Education, 8*. https://doi.org/10.3389/feduc.2023.1120129, 1120129.

Li, B., Li, X., Stoet, G., & Lages, M. (2019). Exploring individual differences in task switching. *Acta Psychologica, 193*, 80–95. https://doi.org/10.1016/j.actpsy.2018.12.010

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(2), 151–173. https://doi.org/10.1207/S15328007SEM0902_1

Loesche, P., Wiley, J., & Hasselhorn, M. (2015). How knowing the rules affects solving the raven advanced progressive matrices test. *Intelligence, 48*, 58–75. https://doi.org/10.1016/j.intell.2014.10.004

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*(2), 130–149.

Marshalek, B., Lohman, D. F., & Snow, R. E. (1983). The complexity continuum in the radex and hierarchical models of intelligence. *Intelligence, 7*(2), 107–127. https://doi.org/10.1016/0160-2896(83)90023-5

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

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Nicolay, B., Krieger, F., & Greiff, S. (2023). Interdisciplinary frontiers: Computer-based process data analysis in educational measurement. In R. J. Tierney, F. Rizvi, & K. Erkican (Eds.), *International Encyclopedia of education* (4th ed., pp. 417–429). Elsevier. https://doi.org/10.1016/B978-0-12-818630-5.10051-X.

Oberauer, K., Süß, H. M., Wilhelm, O., & Wittman, W. W. (2003). The multiple faces of working memory: Storage, processing, supervision, and coordination. *Intelligence, 31*(2), 167–193. https://doi.org/10.1016/S0160-2896(02)00115-0

Pallentin, V. S., Danner, D., & Rummel, J. (2023). Construction and validation of the HeiQ: An operation-oriented figural matrices test. *Journal of Intelligence, 11*(4), 73. https://doi.org/10.3390/jintelligence11040073

Pedersen, T. L. (2025). *patchwork: The Composer of Plots (R Package Version 1.3.1)*. Comprehensive R Archive Network (CRAN). Available online: https://CRAN.R-project.org/package=patchwork (accessed on 10 December 2025).

Prabhakaran, V., Smith, J. A., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. (1997). Neural substrates of fluid reasoning: An fMRI study of neocortical activation during performance of the raven's progressive matrices test. *Cognitive Psychology, 33*(1), 43–63. https://doi.org/10.1006/cogp.1997.0659

R Core Team. (2025). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. Comprehensive R Archive Network (CRAN). Available online: http://www.R-project.org/ (accessed on 10 December 2025).

Raven, J. C. (1962). *Advanced progressive matrices: Sets I and II* (Revised ed.). H.K: Lewis.

Raven, J. C., Court, J. H., & Raven, J. E. (1996). *Standard progressive matrices*. Oxford Psychologists Press.

Revelle, W. (2025). *psych: Procedures for Personality and Psychological Research (R Package Version 2.5.6)*. Comprehensive R Archive Network (CRAN). Available online: https://CRAN.R-project.org/package=psych (accessed on 10 December 2025).

Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General, 124*(2), 207–231. https://doi.org/10.1037/0096-3445.124.2.207

Rosseel, Y. (2012). Lavaan: An r package for structural equation modeling. *Journal of Statistical Software, 48*(1), 1–36.

Roth, B., Becker, N., Romeyke, S., Schäfer, S., Domnick, F., & Spinath, F. M. (2015). Intelligence and school grades: A meta-analysis. *Intelligence, 53*, 118–137. https://doi.org/10.1016/j.intell.2015.09.002

Salthouse, T. A., Fristoe, N., McGuthry, K. E., & Hambrick, D. Z. (1998). Relation of task switching to speed, age, and fluid intelligence. *Psychology and Aging, 13*(3), 445–461. https://doi.org/10.1016/j.actpsy.2006.11.007

Schmitz, F., & Krämer, R. J. (2023). Task switching: On the relation of cognitive Flexibility with cognitive capacity. *Journal of Intelligence, 11*(4), 68. https://doi.org/10.3390/jintelligence11040068

Schulz-Hardt, S. (2025). Zur Lage der Psychologie. *Psychologische Rundschau, 76*(1), 1–20. https://doi.org/10.1026/0033-3042/a000702

Shipstead, Z., & Engle, R. W. (2013). Interference within the focus of attention: Working memory tasks reflect more than temporary maintenance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(1), 277–289. https://doi.org/10.1037/a0028467

Shipstead, Z., Harrison, T. L., & Engle, R. W. (2016). Working memory capacity and fluid intelligence: Maintenance and disengagement. *Perspectives on Psychological Science, 11*(6), 771–799. https://doi.org/10.1177/1745691616650647

Sparfeldt, J. R., Buch, S. R., Rost, D. H., & Lehmann, G. (2008). Akkuratesse selbstberichteter zensuren. *Psychologie in Erziehung und Unterricht, 55*(1), 68–75.

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology, 15*(1), 72–101.

Stadler, M., Hofer, S., & Greiff, S. (2020). First among equals: Log data indicates ability differences despite equal scores. *Computers in Human Behavior, 111*, Article 106442. https://doi.org/10.1016/j.chb.2020.106442

Stadler, M., Brandl, L., & Greiff, S. (2023). 20 years of interactive tasks in large-scale assessments: Process data as a way towards sustainable change? *Journal of Computer Assisted Learning, 39*(6), 1852–1859. https://doi.org/10.1111/jcal.12847

Sticca, F., Goetz, T., Bieg, M., Hall, N. C., Eberle, F., & Haag, L. (2017). Examining the accuracy of students' self-reported academic grades from a correlational and a discrepancy perspective: Evidence from a longitudinal study. *PLoS One, 12*(11), Article e0187367. https://doi.org/10.1371/journal.pone.0187367

Sudevan, P., & Taylor, D. A. (1987). The cuing and priming of cognitive operations. *Journal of Experimental Psychology: Human Perception and Performance, 13*(1), 89–103. https://doi.org/10.1037/0096-1523.13.1.89

SurveyCircle. (2025). Forschungswebseite SurveyCircle. Available online: https://www.surveycircle.com (accessed on 10 December 2025).

Tivian XI GmbH. (2025). Unipark [Software]. https://www.unipark.com.

UNESCO. (2015). *International Standard Classification of Education. Fields of education and training 2013 (ISCED F 2013). Detailed field descriptions*. UNESCO Institute for Statistics.

Unsworth, N., & Engle, R. W. (2005). Working memory capacity and fluid abilities: Examining the correlation between operation span and raven. *Intelligence, 33*(1), 67–81. https://doi.org/10.1016/j.intell.2004.08.003

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

Wang, Y., Zhou, X., Peng, X., & Hu, X. (2022). Task switching involves working memory: Evidence from neural representation. *Frontiers in Psychology, 13*, Article 1003298. https://doi.org/10.3389/fpsyg.2022.1003298

Weber, D., Krieger, F., Spinath, F. M., Greiff, S., Hissbach, J., & Becker, N. (2023). A log file analysis on the validity of partial solutions in figural matrices tests. *Journal of Intelligence, 11*(2), 37. https://doi.org/10.3390/jintelligence11020037

Weber, D., Koch, M., Spinath, F. M., Krieger, F., & Becker, N. (2025). Log file times as indicators of structured figural matrix processing. *Journal of Intelligence, 13*(6), 63. https://doi.org/10.3390/jintelligence13060063

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer.

Wickham, H. (2023). *Stringr: Simple, consistent wrappers for common string operations (r package version 1.4.0)*. Comprehensive r Archive Network (CRAN). Available online: https://CRAN.R-project.org/package=stringr (accessed on 10 December 2025).

Wickham, H., & Henry, L. (2025). Purr: Functional programming tools (r package version 1.1.0). Comprehensive r archive network (CRAN). Available online https://CRAN.R-project.org/package=purrr.

Wickham, H., François, R., Henry, L., Müller, K., Vaughan, D., & Posit software, and PBC. (2023). Dplyr: A grammar of data manipulation (r package version 1.1.4). Comprehensive r archive network (CRAN). Available online https://CRAN.R-project.org/package=dplyr.

Wiley, J., Jarosz, A. F., Cushen, P. J., & Colflesh, G. J. H. (2011). New rule use drives the relation between working memory capacity and raven's advanced progressive matrices. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(1), 256–263. https://doi.org/10.1037/a0021613

Yehene, E., & Meiran, N. (2007). Is there a general task switching ability? *Acta Psychologica, 126*(3), 169–195. https://doi.org/10.1016/j.actpsy.2006.11.007