



TEL AVIV אוניברסיטת תל אביב
UNIVERSITY

Tel Aviv University

Faculty of Humanities

The Jaime and Joan Constantiner School of Education

Department of Educational Counseling and Special Education

Hierarchical Planning and Hierarchical Representation for Algebraic Expressions

Thesis Submitted for the MA Degree of Humanities

Program in Learning Disabilities

by

Hila Zach

The study was conducted under the guidance of

Dr. Dror Dotan

Submission date: September 2026

Word count: 6,658

Abstract

Human cognition relies on hierarchical organization to represent and handle complex symbolic information across domains such as language, music, and motor behaviour. Mathematical expressions are similarly organized according to hierarchical operator structure, yet little is known about how this structure is represented and used. Specifically, it remains unclear whether hierarchical organization reflects a stable cognitive representation of symbolic structure or a planning strategy that reduces working-memory load.

The present study investigated both hierarchical representation and hierarchical planning in the copying of algebraic expressions. Here, hierarchical representation concerns how operator-defined structure is maintained in working memory, whereas hierarchical planning concerns how that structure is used to organize copying into meaningful units rather than writing symbol by symbol. The study examined whether these forms of hierarchical organization are reflected in behavioural performance and how they differ between children and adults.

Twenty-five adults and twenty-seven sixth-grade children completed two complementary copying tasks. In the Hidden Copying Task, the expression remained on screen but was occluded, so that participants could consult it as often as they wished but could not look and write at the same time; this task was designed to capture copying strategy and planning. In the Delayed Copying Task, the expression was shown briefly and then removed, so that the complete expression had to be maintained in memory before writing; this task was designed to probe memory representation. Within each task, we systematically manipulated operator structure—that is, the arrangement of multiplication and addition operators—while holding symbol identity and visual complexity constant. Performance was assessed using copying accuracy, error patterns, and inter-token timing.

The results showed different expressions of sensitivity to operator structure across tasks and age groups. In the Delayed Copying Task, children's copying accuracy differed between the two critical operator structures, whereas adults showed little corresponding difference, providing evidence of hierarchy-congruent representation in children. In the Hidden Copying Task, adults' copying accuracy differed between the two critical operator structures, while children did not show a comparable structure-dependent difference, providing evidence of hierarchy-congruent planning in adults. In both tasks, the inter-token gap profiles showed descriptive structure-related differences, but the critical Position $\times$ Condition interactions did not reach statistical significance. In the Hidden Copying Task, the adult timing-error correlation was significant in one critical structure but not the other; however, the direct difference between these correlations was not statistically significant.

Together, these findings suggest that operator structure can influence both the representation maintained in working memory and the planning strategy used during copying. The different patterns across age groups indicate that hierarchy-congruent representation and planning are not expressed identically in children and adults. Future research should clarify how these components develop and whether making operator-defined grouping explicit can support the learning and writing of mathematical expressions.

תקציר

ארגון היררכי הוא אחד מעקרונות היסוד של הקוגניציה האנושית, ומאפשר לייצג ולהתמודד עם מידע סימבולי מורכב בתחומים כגון שפה, מוזיקה ותכנון מוטורי. גם ביטויים מתמטיים מאורגנים על פי מבנה היררכי הנקבע באמצעות קדימות האופרטורים, אולם עדיין לא ברור כיצד מבנה זה מיוצג וכיצד נעשה בו שימוש. בפרט, לא ברור האם הארגון ההיררכי משקף ייצוג קוגניטיבי יציב של המבנה הסימבולי או אסטרטגיית תכנון המפחיתה את העומס על זיכרון העבודה.

המחקר הנוכחי בחן הן ייצוג היררכי והן תכנון היררכי בהעתקה של ביטויים אלגבריים. ייצוג היררכי מתייחס לאופן שבו המבנה המוגדר על ידי האופרטורים נשמר בזיכרון העבודה, ואילו תכנון היררכי מתייחס לאופן שבו מבנה זה משמש לארגון ההעתקה ליחידות בעלות משמעות מבנית במקום כתיבה של סימן אחר סימן. המחקר בחן האם צורות אלה של ארגון היררכי באות לידי ביטוי בביצוע וכיצד הן שונות בין ילדים למבוגרים.

במחקר השתתפו 25 מבוגרים ו-27 תלמידי כיתה ו'. המשתתפים ביצעו שתי מטלות העתקה משלימות. במטלת Hidden Copying הופיע הביטוי על המסך אך הוסתר, כך שהמשתתפים יכלו להציץ בו כמה פעמים שרצו אך לא יכלו להסתכל ולכתוב בו-זמנית; מטלה זו נועדה לבחון את אסטרטגיית ההעתקה והתכנון. במטלת Delayed Copying הוצג הביטוי לזמן קצר ולאחר מכן נעלם, כך שהיה צורך לשמור את הביטוי המלא בזיכרון לפני הכתיבה; מטלה זו נועדה לבחון את הייצוג בזיכרון. בכל מטלה שינינו באופן שיטתי את מבנה האופרטורים, כלומר את סידור פעולות הכפל והחיבור, תוך שמירה על זהות הסמלים והמורכבות החזותית. הביצוע הוערך באמצעות דיוק ההעתקה, דפוסי השגיאות ופערי הזמן בין הטוקנים.

הממצאים הראו ביטויים שונים של רגישות למבנה האופרטורים בין המטלות ובין קבוצות הגיל. במטלת Delayed Copying נמצא הבדל בדיוק ההעתקה של הילדים בין שני המבנים הקריטיים, בעוד שבקרב המבוגרים נמצא הבדל קטן בלבד, ממצא המספק עדות לייצוג היררכי תואם-מבנה אצל הילדים. במטלת Hidden Copying נמצא אצל המבוגרים הבדל בדיוק ההעתקה בין שני המבנים הקריטיים, ואילו אצל הילדים לא נמצא הבדל מקביל, ממצא המספק עדות לתכנון היררכי תואם-מבנה אצל המבוגרים. בשתי המטלות נראו הבדלים תיאוריים בדפוסי פערי הזמן בין הטוקנים, אך האינטראקציות הקריטיות בין עמדה לתנאי לא הגיעו למובהקות סטטיסטית. במטלת Hidden Copying המתאם בין זמן ההשהיה לשיעור השגיאות היה מובהק אצל המבוגרים באחד המבנים הקריטיים אך לא באחר; עם זאת, ההבדל הישיר בין שני המתאמים לא היה מובהק סטטיסטית.

ממצאים אלה מצביעים על כך שמבנה האופרטורים יכול להשפיע הן על הייצוג הנשמר בזיכרון העבודה והן על אסטרטגיית התכנון המשמשת במהלך ההעתקה. הדפוסים השונים בין קבוצות הגיל מצביעים על כך שייצוג היררכי ותכנון היררכי אינם מתבטאים באופן זהה אצל ילדים ומבוגרים. מחקר עתידי נדרש לבחון כיצד רכיבים אלה מתפתחים והאם הדגשה מפורשת של הקיבוץ המוגדר על ידי האופרטורים יכולה לתמוך בלמידה ובכתיבה של ביטויים מתמטיים.

Table of Contents

Abstract	2
תקציר	3
Table of Contents	4
1 Introduction.....	5
1.1 Hierarchical Cognition: A Fundamental Principle of Human Thought.....	5
1.2 Hierarchical Structure in Mathematics.....	5
1.3 Mental Representation, Working Memory, and Chunking.....	6
1.4 Investigating Hierarchical Representation Through Copying Tasks.....	8
1.5 Development of Hierarchical Representation.....	8
1.6 Research Questions.....	9
2 Method.....	9
2.1 Participants.....	9
2.2 The Copying Tasks.....	9
2.3 Data Coding.....	10
2.4 Statistical Analyses.....	11
3 Results.....	11
3.1 The Delayed Copying Task - A Measure of Hierarchical Representation.....	11
3.1.1 Accuracy Shows Hierarchy-Congruent Representation in Children.....	11
3.1.2 Inter-Token Gaps Show a Descriptive Trend Toward Hierarchy-Congruent Representation.....	12
3.1.3 Summary of the Delayed Copying Task.....	12
3.2 The Hidden Copying Task.....	13
3.2.1 Accuracy Shows Hierarchy-Congruent Planning in Adults.....	13
3.2.2 Inter-Token Gaps Show a Descriptive Trend Toward Hierarchy-Congruent Planning.....	13
3.2.3 Error Types Reveal Different Patterns of Production Errors Across Age Groups.....	14
3.2.4 Summary of the Hidden Copying Task.....	15
4 Discussion.....	15
4.1 Evidence for a Hierarchical Mental Representation.....	15
4.2 Evidence for Hierarchical Planning.....	16
4.3 Working Memory and Hierarchical Organisation.....	16
4.4 Developmental Differences.....	16
4.5 Theoretical and Educational Implications.....	17
4.6 Open Questions.....	17
4.7 Conclusion.....	18
References	19
Appendix A. Z-Scored Inter-Token Gap Analyses	21

1 Introduction

1.1 Hierarchical Cognition: A Fundamental Principle of Human Thought

One of the defining characteristics of human cognition is the ability to organize individual elements into meaningful hierarchical structures (Fitch & Martins, 2014; Harrison & Stiles, 2009; Van de Cavey & Hartsuiker, 2016; Martins, 2025). Rather than processing information as isolated or sequential units, the cognitive system groups related elements into nested representations to support efficient reasoning, memory, and action. This capacity has been documented across numerous cognitive domains. In language, linguistic theories propose that sentences are represented as nested hierarchical syntactic structures rather than as simple linear sequences of words (Chomsky, 1957, 1965). Similar hierarchical representational principles have been identified in motor behaviour, where complex actions are decomposed into nested subgoals (Lashley, 1951; Rosenbaum et al., 2007); in music, where listeners organize notes into larger rhythmic and harmonic structures (Lerdahl & Jackendoff, 1983); and in visual cognition, where perceptual grouping enables observers to integrate local elements into coherent global forms (Palmer, 1999). The consistent findings across these domains suggest that hierarchy constitutes a general principle of human cognition rather than a property unique to any single symbolic system.

Hierarchical representations allow multiple lower-level elements to be organised into larger functional units, or chunks. In working memory, chunking can reduce the number of independent elements that must be actively maintained (Miller, 1956; Gobet et al., 2001; Chekaf et al., 2016), while recurring groupings can also become stored units in long-term memory (Gobet et al., 2001; Musfeld et al., 2024). In the present context, the relevant question is whether the formal relations among the elements of an algebraic expression guide such grouping.

1.2 Hierarchical Structure in Mathematics

Although hierarchical organization has been extensively investigated in language and other symbolic systems, considerably less is known about how hierarchical representations are constructed during mathematical processing. Arithmetic expressions, e.g., $4+3\times 2$, provide an especially valuable domain for addressing this question. Unlike natural language, whose structure is inferred from grammatical rules and contextual information, algebraic expressions are governed by formal syntactic principles, such as operator precedence, that explicitly specify how the symbols are grouped. They therefore offer a unique opportunity to examine how hierarchical symbolic relationships are mentally represented, independently of many of the semantic and contextual factors that characterize natural language. They also offer the uncommon opportunity to examine hierarchical processing in a domain that is different from natural language, and yet, like natural language, contains stimuli with clear and concrete meanings.

The structure of a simple arithmetic expression can be specified by parentheses or, when parentheses are absent, by operator precedence. For example, $3 \times 4 + 5 \times 6$ is interpreted as $(3 \times 4) + (5 \times 6)$: the two multiplications form two equivalent constituents that are subsequently combined by addition, producing a balanced 2-2 structure. In contrast, $3 + 4 \times 5 + 6$ is interpreted as $3 + (4 \times 5) + 6$: a single

multiplication constituent is embedded between two additions, producing a 1-2-1 structure. Thus, expressions containing the same number of operands and operators can impose different hierarchical groupings even without explicit parentheses.

Cognitively, this means that reading such an expression cannot consist merely of recognising its individual symbols. It requires constructing an internal representation of the relations among them, and the two expressions above, despite containing the same number of operands and operators and therefore the same amount of symbolic information, call for different representations.

Previous research has demonstrated that people are sensitive to these hierarchical structures during mathematical processing. Rather than reading algebraic expressions as simple linear strings, individuals appear to organise expressions into meaningful structural units that correspond to operator precedence rules. Behavioural studies have shown that operator structure influences reading times, eye movements, calculation strategies, and error patterns, suggesting that hierarchical organisation affects multiple stages of mathematical cognition (Kirshner, 1989; Landy & Goldstone, 2007; Scheepers et al., 2011; Marghetis et al., 2016). Importantly, Scheepers et al. (2011) manipulated grouping using explicit parentheses, whereas the present study examines expressions without parentheses, so the relevant grouping must be inferred from operator precedence.

Two explanations are relevant to the present study. One possibility is that readers consistently construct internal hierarchical representations that mirror the formal mathematical structure of the expression. Another is that hierarchical cognitive structure becomes relevant primarily when executing a calculation, reflecting learned procedures for applying operator precedence. These explanations are not mutually exclusive, but they make different predictions about whether sensitivity to structure should appear when participants reproduce expressions without calculating them.

Arbitrating between these explanations is challenging because most experimental tasks confound hierarchical representation with overt mathematical computation that presumably relies on learned strategies. During equation solving, for example, participants rely on calculation procedures, prior mathematical knowledge, or strategic problem-solving processes that obscure the underlying representational mechanisms. Consequently, improved performance on hierarchically organised expressions does not necessarily demonstrate that the expressions themselves are mentally represented hierarchically.

To distinguish these possibilities, it is useful to examine situations in which participants encode and reproduce expressions without calculating them. Copying preserves the hierarchical relationships among the symbols while reducing the contribution of calculation procedures. The objective here is therefore not to determine whether participants can solve the expressions correctly, but to examine how their structural organisation is represented during encoding and written production.

1.3 Mental Representation, Working Memory, and Chunking

A typical processing pathway of an arithmetic expression presumably consists of several processing stages: visual processing of the expression, maintaining it in memory, and—if the task requires solving the expression—performing calculations. In each of these stages, the processing could be done by different cognitive processes, and could additionally be affected by learned strategies.

The ability to perceive the hierarchical structure of an algebraic expression does not, by itself, explain how that structure is mentally represented. Once visual input disappears, successful performance depends on the cognitive processes responsible for maintaining, organizing, and reproducing symbolic information. Consequently, understanding how algebraic expressions are mentally represented requires consideration of the mechanisms through which hierarchical information is encoded, stored, and transformed into written output.

Working memory plays a critical role in this process. It enables information to remain temporarily available while it is manipulated, integrated, and transformed into goal-directed actions (Baddeley & Hitch, 1974; Baddeley, 2000). During the reproduction of an algebraic expression, participants must maintain the symbolic information while simultaneously planning and executing the motor actions required for writing. Because working memory has limited capacity, efficient performance depends on reducing cognitive load through the construction of organized representations.

One of the principal mechanisms through which this reduction is achieved is chunking. Chunking refers to the process of grouping multiple individual elements into larger, meaningful units that can be processed as single cognitive entities (Miller, 1956; Gobet et al., 2001). By combining several related elements into one integrated representation, chunking reduces the number of independent units that must be maintained simultaneously in working memory. This process has been demonstrated across numerous domains, including language, chess expertise, music, motor behaviour, and mathematical cognition, suggesting that it reflects a general cognitive mechanism rather than a domain-specific strategy.

Within algebraic expressions, chunking is expected to follow the hierarchical relationships specified by operator precedence. For example, in the expression

$$(a \times a) + (a \times a),$$

the two multiplication operations naturally form two higher-order units before they are combined through addition. Instead of maintaining seven independent symbols, the cognitive system may represent the expression as two integrated sub-expressions connected by a higher-level operation. Such an organization substantially reduces working-memory demands while preserving the mathematical relationships among the components.

Importantly, chunking should not be understood as merely a memory aid. Rather, it may reflect the underlying structure of the mental representation itself. If hierarchical chunking forms part of the representation, behavioural measures should reveal its influence even in situations where explicit calculation is unnecessary. Conversely, if chunking reflects only a consciously adopted mnemonic strategy, its effects should depend strongly on task demands and may disappear when memory requirements are reduced.

This distinction between representation and strategy is central to the theoretical motivation of the present study. A mental representation refers to the cognitive organisation of information itself—the way symbolic relationships are internally encoded and maintained. A strategy, in contrast, refers to the deliberate procedures individuals adopt to improve performance under particular task conditions. Although the two concepts are closely related, they are not equivalent. Individuals may employ strategies that operate on existing representations, but the presence of a strategy does not necessarily reveal the structure of the representation itself.

1.4 Investigating Hierarchical Representation Through Copying Tasks

Mental representations cannot be observed directly, so they must be inferred from behaviour using tasks that minimise processes unrelated to representation, such as explicit calculation. Copying paradigms are useful for this purpose because participants reproduce expressions that are already complete. The reproduced expression preserves every operator, operand, and hierarchical relationship of the original, allowing accuracy, error patterns, and production timing to be related to the structural organisation of the expression.

The present study used two copying tasks that differed in participants' access to the expression while writing. In the Hidden Copying Task, the expression remained available but was occluded whenever the participant wrote, so participants could consult it as often as they wished but could not look and write at the same time. In the Delayed Copying Task, the expression was shown briefly and then removed, so it had to be reproduced from memory.

Comparing the two tasks helps distinguish representation from copying strategy. In the Delayed Copying Task, participants must maintain the complete expression in memory before reproducing it, so performance provides evidence about the representation maintained in memory. In the Hidden Copying Task, participants can return to the expression repeatedly and determine how much of it to encode before each production segment. How they divide the expression across these segments therefore provides information about their copying strategy, while still depending on the representation available to support each segment.

1.5 Development of Hierarchical Representation

Comparing children and adults provides two points along the development of hierarchical representation. The present study focuses specifically on whether the syntactic structure of algebraic expressions is represented differently across these age groups, rather than on general developmental differences in executive abilities.

Previous research provides evidence for developmental differences in hierarchical processing. Martins, Laaha, Freiberger, Choi and Fitch (2014) found developmental differences in children's representation of hierarchical visual structures. In the mathematical domain, Braithwaite, Goldstone, van der Maas and Landy (2016) showed that children's use of operator precedence is influenced by perceptual structure, and Marghetis, Landy and Goldstone (2016) found sensitivity to constituent boundaries among participants who had mastered operator precedence. However, these studies do not show whether children and adults differ in the internal representation and production of algebraic structure when no calculation is required. The present study addresses this gap by comparing children and adults in copying tasks that preserve operator structure while removing the need to solve the expressions.

Both age groups reproduced the same expressions without calculating them, allowing age-related differences in sensitivity to operator structure to be examined during representation and production.

1.6 Research Questions

The present study addresses the following three research questions:

1. **Does the hierarchical operator structure of algebraic expressions influence the accuracy with which they are reproduced?** Prediction: expressions with balanced operator structures will be reproduced more accurately than expressions with a less balanced hierarchical organisation.
2. **Do production timing measures provide behavioural evidence for hierarchical planning during written reproduction?** Prediction: pauses will be longer at the locations corresponding to higher-order structural boundaries within the expression.
3. **Are these effects influenced by developmental stage and by the availability of visual information during production?** Prediction: adults will reproduce expressions more accurately than children, reflecting more efficient hierarchical organisation of symbolic information; and structural effects will be more pronounced in the Delayed than in the Hidden Copying Task, because higher working-memory demands amplify the influence of the internal representation.

2 Method

2.1 Participants

Participants were initially 29 sixth-grade children (17 girls, 12 boys; mean age 11.8 years, $SD = 0.4$, range 11.1–13.0) and 31 adults (16 women, 15 men; mean age 34.5 years, $SD = 10.1$, range 22–57); the final study sample comprised 27 children and 25 adults, as reported in the Abstract. Sixth-graders were chosen because they had already been taught arithmetic and operator precedence. The adults were compensated for participation.

Adults completed a background questionnaire. Eleven had studied mathematics at the 3-unit matriculation level, eight at 4 units, and eleven at 5 units. Educational attainment ranged from high school to doctorate. One adult reported a diagnosed learning or attention disorder and two were unsure. Self-rated mathematics anxiety, on a 1–5 scale, averaged 2.8 ($SD = 1.4$). Before the copying tasks, children completed an order-of-operations worksheet and a multiplication-table worksheet.

Adults gave written informed consent. For children, written consent was obtained from a parent or legal guardian and verbal assent from the child. All participants had normal or corrected-to-normal vision. The study was approved by Tel Aviv University's Institutional Review Board. Participation was voluntary.

2.2 The Copying Tasks

In both tasks, participants copied arithmetic expressions with a pen on blank paper placed on a digitizing tablet. The pen trajectory was recorded using WriTracker (Dotan et al., 2020). Expressions were presented on a separate screen.

Both tasks were administered individually in a quiet room within a single session, lasting approximately one hour for children and 45 minutes for adults. The Hidden Copying Task was always administered first, followed by the Delayed Copying Task. Each task comprised 48 experimental trials. All participants completed both tasks, except for those who declined to continue or were unable to do so; these cases are described below in the Data Coding section.

Hidden Copying Task. The expression remained on the screen but was occluded by a sheet of cardboard whenever the participant wrote. Participants could look behind the cardboard as often and for

as long as they wished, but could not look and write at the same time. They were told that there was no time limit, that accuracy was important, and that they could not check their response after finishing it. There were no practice trials.

Delayed Copying Task. The expression appeared on the screen for one second and then disappeared. Participants wrote as much of the expression as they could remember and were instructed not to begin writing while it was still on the screen. The task was preceded by two practice trials, which were excluded from the analyses.

Stimuli. Each stimulus was an arithmetic expression containing 4 operands and 3 operators. There were 3 conditions, with 16 stimuli in each. One critical condition was 2-2, with the operator structure $\times + \times$, e.g., $8 \times 9 + 6 \times 2$. Operator precedence groups this expression into two multiplication constituents, $(8 \times 9) + (6 \times 2)$. The second critical condition was 1-2-1, with the operator structure $+ \times +$, e.g., $8 + 9 \times 6 + 2$, in which one multiplication constituent is embedded between two additions. A third, baseline condition used only additions (Fixed-Operator condition). All conditions used the same 16 operand quadruples. The 48 stimuli were presented in a single mixed block in the same random order for all participants. Consecutive trials never had the same condition or operand quadruple. The digits 0 and 1 were not used, and no operand was repeated within an expression. In the Hidden Copying Task, all operands were single digits. In the Delayed Copying Task, operands were single digits for all participants except one adult participant, who received two-digit operands.

2.3 Data Coding

Pen trajectories were segmented and labelled with the WriTracker encoder. A response was scored as correct when the written expression matched the target in the identity and order of every operand and operator.

Errors were classified as incorrect-operand errors, incorrect-operator errors, or other errors. Error trials were not included in the inter-token-gap analyses. Trials containing self-corrections, erasures, stray marks, superfluous strokes, or a later return to an earlier position to add a symbol were also excluded from the inter-token-gap analyses.

For the Delayed Copying Task, data were available for 26 children and 25 adults. Participants 113, 123, and 125 were excluded because they had fewer than five usable correct trials overall or in at least one condition. The final Delayed Copying analyses therefore included 48 participants (23 children and 25 adults).

For the timing analyses, a token was defined as an operator, a one-digit operand, or a two-digit operand. An inter-token temporal gap was defined as the time elapsed from lifting the pen from the paper at the end of one token until placing it back on the paper to begin writing the next token. Gaps between the two digits of a two-digit operand and the gap preceding the first operand in each trial were not analysed.

2.4 Statistical Analyses

Inter-token temporal gaps were analysed using raw gap durations. For the critical hypothesis, pre-token delay was entered as the dependent variable in a linear mixed model, with participant as a random

factor and Condition, Position, and their interaction as factors. Analyses were conducted separately for each age group. The critical effect was the Position $\times$ Condition interaction. The same focused analyses after z-scoring gap durations within participant are reported as a supplementary robustness check in Appendix A.

3 Results

3.1 The Delayed Copying Task - A Measure of Hierarchical Representation

The Delayed Copying Task examined participants' representation of algebraic expressions when the complete expression had to be maintained in memory before production. The expression was presented briefly and then disappeared, requiring participants to maintain an internal representation of its elements and structure while reproducing it. The task therefore provides a measure of how the expression is represented in memory. Our analyses below focused on the two critical operator structures, 2-2 and 1-2-1. The Fixed-Operator condition is reported merely as a baseline condition, which is largely independent of our research questions.

3.1.1 Accuracy Shows Hierarchy-Congruent Representation in Children

Children were more accurate in the 2-2 condition ($M = 62.5\%$, $SD = 26.2\%$) than in the 1-2-1 condition ($M = 45.0\%$, $SD = 24.2\%$), $t(22) = 3.98$, $p < .001$. We refer to this pattern as hierarchy-congruent representation: reproduction is more accurate when the operator structure supports the balanced 2-2 organisation than when it supports the asymmetrical 1-2-1 organisation.

We propose that the operator structure of the stimulus affected the representation constructed in memory. In the 2-2 condition, the $\times + \times$ operator sequence encouraged a representation consisting of two chunks, $(a \times b)$ and $(c \times d)$, which were then linked by addition. In the 1-2-1 condition, the $+ \times +$ sequence encouraged a 3-chunk representation, with one central multiplication constituent flanked by two separate operands: $(a) + (b \times c) + (d)$. Maintaining two balanced chunks presumably places lower demands on memory than maintaining the 1-2-1 organization (Chekaf et al., 2016), which could account for the lower error rate in the 2-2 condition. Thus, the accuracy difference between the conditions suggests that children's representation reflected the operator structure of the stimulus rather than only the linear sequence of its elements.

Adults showed no comparable accuracy difference: mean accuracy was 75.0% ($SD = 14.2\%$) in the 2-2 condition and 71.2% ($SD = 20.1\%$) in the 1-2-1 condition, $t(24) = 0.83$, $p = .42$. Accuracy therefore provided no corresponding evidence of hierarchy-congruent representation in adults. Possible explanations for why this effect was observed in children but not in adults are considered in the Discussion.

3.1.2 Inter-Token Gaps Show a Descriptive Trend Toward Hierarchy-Congruent Representation

If the memory representation is hierarchy-congruent, the relative size of the gaps before tokens 4 and 6 should reverse between the two critical conditions. In the 2-2 sequence ($\times + \times$), the gap before token 4 precedes the central addition, which we assume may serve as a chunk boundary, whereas the gap before token 6 precedes the final multiplication, so pre-4 is predicted to be longer than pre-6. In the 1-2-1

sequence (+ × +), these structural roles reverse, because the chunk boundaries are expected to be at each + operator, so pre-6 is predicted to be longer than pre-4. The critical test is therefore the Position (pre-4 vs. pre-6) × Condition (2-2 vs. 1-2-1) interaction.

The raw inter-token temporal gaps showed a descriptive pattern in the predicted direction in children (Figure 1), but the Position × Condition interaction was not significant: in a linear mixed model with the pre-token delay as the dependent variable, the participant as the random factor, and Condition, Position, and their interaction as factors, the interaction term showed $F(1, 22) = 2.33, p = .14$. Adults showed nearly parallel profiles, and the corresponding model also showed no significant Position × Condition interaction, $F(1, 24) = 0.003, p = .96$. Thus, the raw timing data provide no statistical evidence for hierarchy-congruent representation. Essentially the same pattern was obtained after z-scoring gap durations within participant; this supplementary analysis, which controls for individual differences in overall timing scale, is reported in Appendix A.

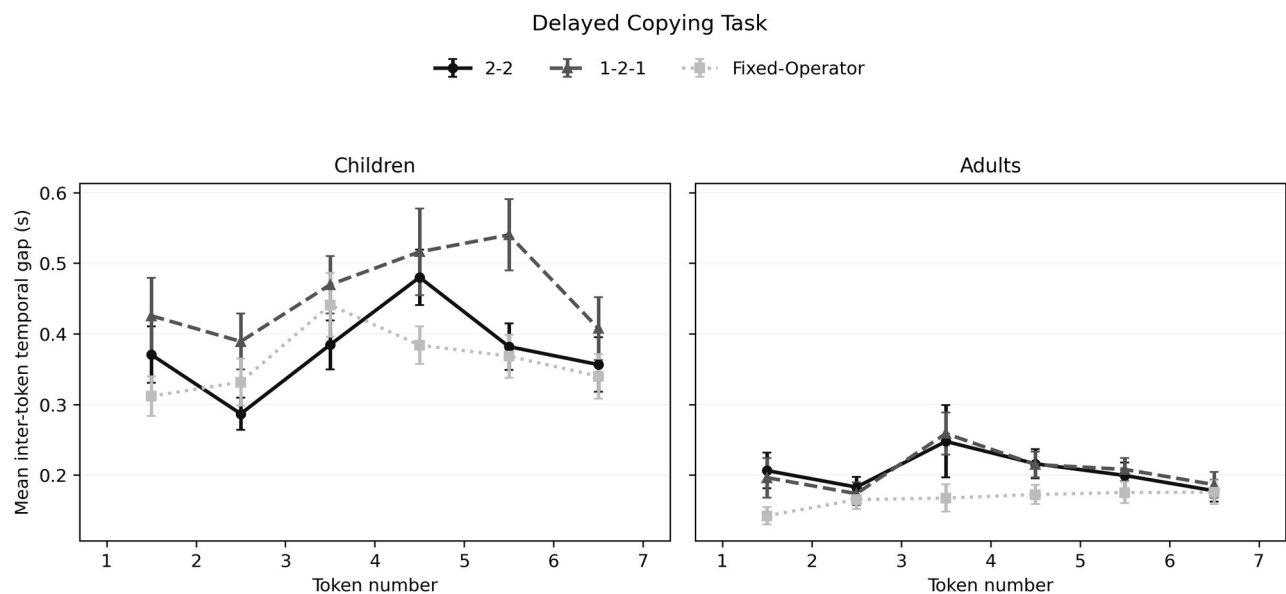


Figure 1. Delayed Copying Task. Mean raw inter-token temporal gaps (s) by token position for children and adults. The 2-2 and 1-2-1 conditions are the critical comparison; the Fixed-Operator condition is shown as a lighter baseline. In children, the 1-2-1 profile was higher at pre-6 than at pre-4, while the 2-2 profile was nearly flat across those two positions; the critical Position × Condition interaction was not significant. Adults showed nearly parallel 2-2 and 1-2-1 profiles. Error bars represent ± 1 SEM.

3.1.3 Summary of the Delayed Copying Task

The main theoretically relevant finding from the Delayed Copying Task was hierarchy-congruent representation in children's accuracy: children reproduced 2-2 expressions more accurately than 1-2-1 expressions. The timing data showed a descriptive pre-4/pre-6 pattern in the predicted direction in children, but the critical interaction was not significant; adults showed neither the accuracy difference nor the predicted timing interaction.

3.2 The Hidden Copying Task

The Hidden Copying Task examined how participants organised production when they could return to the expression repeatedly but could not look at it while writing. Unlike the Delayed Copying Task, participants could therefore determine how much of the expression to memorize before each production segment. The task consequently provides information about how operator structure influences the planning of the copying process.

3.2.1 Accuracy Shows Hierarchy-Congruent Planning in Adults

Adults were more accurate in the 2-2 condition ($M = 80.9\%$, $SD = 11.1\%$) than in the 1-2-1 condition ($M = 74.1\%$, $SD = 18.3\%$), $t(24) = 2.68$, $p = .01$. We refer to this pattern as hierarchy-congruent planning: performance is better when the operator structure supports division into two balanced production units, presumably because the participant planned the division this way, than when it supports the asymmetrical 1-2-1 organisation.

Children did not show a comparable hierarchy-congruent planning pattern: mean accuracy was 74.3% ($SD = 21.2\%$) in the 2-2 condition and 75.9% ($SD = 22.3\%$) in the 1-2-1 condition, $t(14) = -0.35$, $p = .73$. Thus, Hidden-task accuracy provided evidence of hierarchy-congruent planning in adults but not in children.

3.2.2 Inter-Token Gaps Show a Descriptive Trend Toward Hierarchy-Congruent Planning

The critical timing prediction was the same pre-4/pre-6 crossover described for the Delayed task: 2-2 should show a relatively longer pre-4 gap, whereas 1-2-1 should show a relatively longer pre-6 gap. The raw Hidden-task means were descriptively in this direction in both age groups (Figure 3). In children, the Position $\times$ Condition interaction was not significant: in a linear mixed model with the pre-token delay as the dependent variable, the participant as the random factor, and Condition, Position, and their interaction as factors, the interaction term showed $F(1, 13) = 2.76$, $p = .12$. In adults, the corresponding model also showed no significant Position $\times$ Condition interaction, $F(1, 20) = 0.16$, $p = .69$. Thus, there is no statistical evidence for hierarchy-congruent planning in the inter-token-gap analysis; the crossover is descriptive only. The z-scored supplementary analysis, which showed essentially the same results, is reported in Appendix A.

After participant 213 was excluded as an outlier, adults still showed a positive association between mean pre-production delay and error rate in the 1-2-1 condition, $r = .54$, $p = .007$, whereas the corresponding correlation in the 2-2 condition was not significant, $r = -.13$, $p = .54$. The direct difference between these two dependent correlations did not reach significance in a one-tailed paired condition-label permutation test, $p = .06$. Thus, the condition-wise correlations differ descriptively, but they do not provide statistically conclusive evidence that the timing-error relationship is stronger in 1-2-1 than in 2-2.

Hidden Copying Task - Adults (participant 213 excluded)

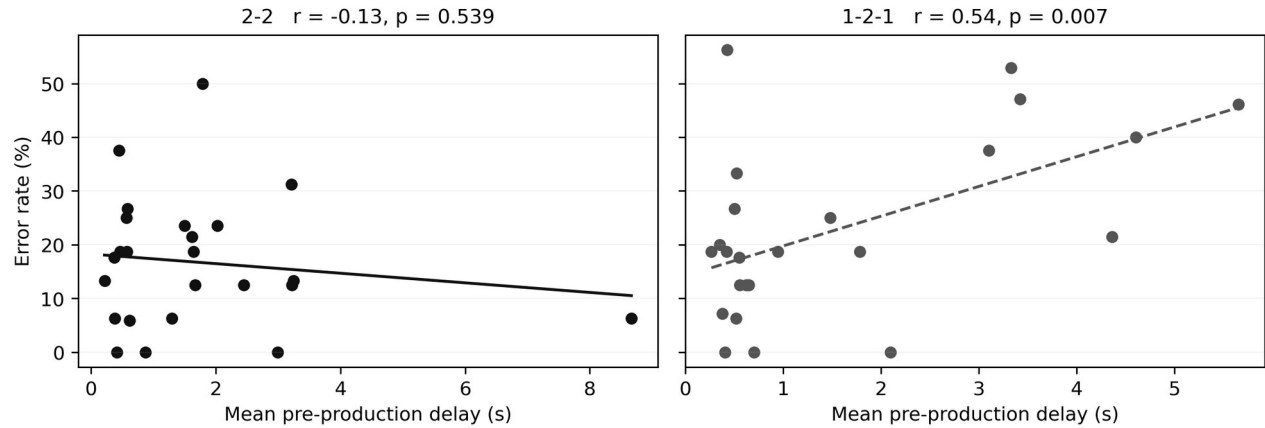


Figure 2. Hidden Copying Task. Mean pre-production delay and error rate for adults in the two critical conditions after excluding participant 213. The correlation was positive in 1-2-1 ($r = .54$, $p = .007$) and non-significant in 2-2 ($r = -.13$, $p = .54$); the direct difference between correlations did not reach significance (one-tailed paired condition-label permutation test, $p = .06$).

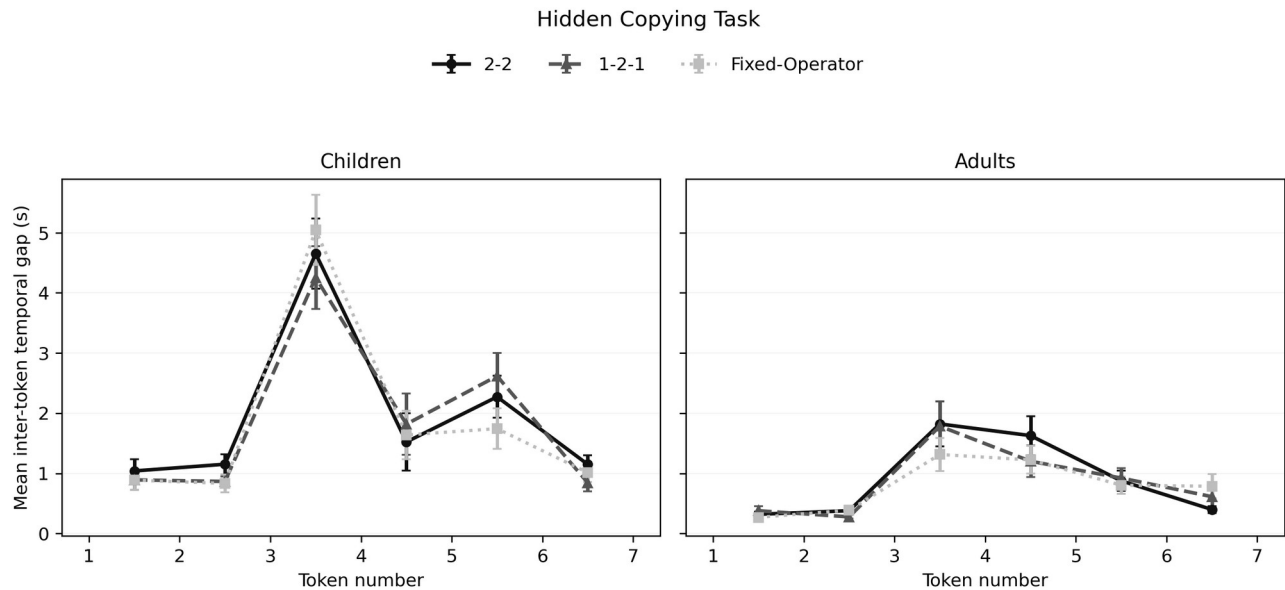


Figure 3. Hidden Copying Task. Mean raw inter-token temporal gaps (s) by token position for children and adults. The 2-2 and 1-2-1 conditions are the critical comparison; Fixed-Operator is shown as a lighter baseline. At pre-4, the 2-2 gap was descriptively longer than 1-2-1, whereas at pre-6 the ordering reversed in both groups, as predicted; neither Position $\times$ Condition interaction was statistically significant. Error bars represent ± 1 SEM.

3.2.3 Error Types Reveal Different Patterns of Production Errors Across Age Groups

The error-type differences were modest. Across Hidden-task trials, incorrect-operand errors occurred on 6.3% of adult trials and 10.7% of child trials; operator-related errors (WrongSign or WrongSignNumber) occurred on 5.5% of adult trials and 5.4% of child trials; and PartialResponse errors occurred on 3.8% of

adult trials and 4.6% of child trials. These descriptive differences do not by themselves establish hierarchy-congruent planning, and the present data do not justify a more specific cognitive explanation.

3.2.4 Summary of the Hidden Copying Task

The main evidence for hierarchy-congruent planning in the Hidden Copying Task came from adults' accuracy: adults reproduced 2-2 expressions more accurately than 1-2-1 expressions, whereas children did not show this difference. The pre-4/pre-6 timing profiles showed the predicted descriptive crossover in both age groups, but neither Position $\times$ Condition interaction was significant, so the gap data provide no statistical evidence for hierarchy-congruent planning. After participant 213 was excluded, the adult timing-error correlation was significant in 1-2-1 but not in 2-2; however, the direct difference between the two correlations did not reach significance.

Error-type differences between age groups were modest and are treated as descriptive; these too suggest a developmental effect, whose origin may be explored in future works.

4 Discussion

The present study investigated whether the hierarchical operator structure of algebraic expressions influences their representation and production, and whether these effects differ between children and adults. The two copying tasks provided complementary information: the Delayed Copying Task required the complete expression to be maintained in memory, whereas the Hidden Copying Task allowed repeated consultation of the expression and therefore provided information about how production was divided and planned.

Overall, hierarchy-congruent effects were expressed differently across tasks and age groups. In the Delayed Copying Task, the clearest accuracy effect was observed in children. In the Hidden Copying Task, the clearest accuracy effect was observed in adults. The timing profiles in both tasks showed descriptive differences between the two critical structures, although the critical Position $\times$ Condition effects were not statistically significant.

Importantly, these effects emerged during reproduction rather than calculation. Participants were not required to evaluate the expressions, making it unlikely that the observed patterns can be attributed solely to calculation strategies. Instead, the findings concern how symbolic expressions are represented and organised during production.

4.1 Evidence for a Hierarchical Mental Representation

The Delayed Copying Task provides the most direct evidence about the representation maintained in memory. Children's higher accuracy in the 2-2 than in the 1-2-1 condition provides evidence of hierarchy-congruent representation. Under the hierarchical-representation account, the balanced 2-2 structure supports two parallel chunks, whereas the 1-2-1 structure supports a less balanced organisation. Adults showed little corresponding difference in accuracy. Possible explanations for this age difference are considered below.

The timing results were compatible with this interpretation at a descriptive level, particularly in children, but the critical Position $\times$ Condition interaction between the 2-2 and 1-2-1 structures did not reach significance. The timing data therefore do not provide statistical evidence for hierarchy-congruent representation; the observed pattern is descriptive only.

An important unresolved question concerns why hierarchical organisation arises. The present design cannot distinguish whether the observed organisation primarily reflects general cognitive constraints, properties of mathematical notation, or an interaction between the two. Future experiments that dissociate mathematically meaningful grouping from memory-efficient grouping would help discriminate between these possibilities.

The absence of a comparable Delayed accuracy effect in adults may have several explanations. Adults may maintain the expressions accurately enough that differences between the two structures are less visible in this measure, or they may rely on a representation or strategy that reduces the accuracy cost of the less balanced structure. The present data do not distinguish between these possibilities.

4.2 Evidence for Hierarchical Planning

The Hidden Copying Task provides complementary evidence about hierarchy-congruent planning during production. Adults were less accurate in the 1-2-1 than in the 2-2 condition, indicating that their accuracy showed hierarchy-congruent planning. After participant 213 was excluded, the timing-error correlation was significant in the 1-2-1 condition but not in the 2-2 condition; however, the direct difference between the two correlations was not statistically significant. The timing-error pattern is therefore descriptive and does not by itself establish a structure-specific effect.

Children did not show a comparable structure-dependent accuracy difference in the Hidden Copying Task. Their inter-token gap profiles, like those of adults, showed descriptive differences between the critical structures at theoretically relevant positions, but the critical interactions were not statistically significant.

Thus, the Hidden Copying findings provide evidence for hierarchy-congruent planning from adults' accuracy. The timing profiles in both age groups showed descriptive trends, but no statistically significant hierarchy-congruent timing effect.

4.3 Working Memory and Hierarchical Organisation

The comparison between the Hidden and Delayed Copying Tasks helps distinguish representation from production strategy. In the Delayed Copying Task, the complete expression had to be maintained in memory before production, whereas in the Hidden Copying Task participants could return to the stimulus repeatedly and choose how much to encode before each production segment.

Hierarchy-congruent effects were observed in both tasks, but in different measures and age groups. Children's Delayed accuracy provides evidence that operator structure affected the representation maintained in memory, whereas adults' Hidden accuracy indicates that operator structure also affected the organisation of production. The Hidden timing-error correlations differed descriptively between the two critical conditions, but their direct difference was not statistically significant.

4.4 Developmental Differences

A second aim of this study was to compare children and adults. The results do not support a simple developmental pattern in which one age group is uniformly more strongly hierarchy-congruent than the other. Instead, hierarchy-congruent effects appeared in different measures across the two tasks.

Children showed the clearest hierarchy-congruent representation effect in the Delayed Copying Task, whereas adults showed the clearest accuracy effect in the Hidden Copying Task. The timing profiles in both groups showed descriptive hierarchy-congruent trends, but the critical interactions were not statistically

significant. The developmental difference therefore concerns how hierarchy-congruent effects were expressed across tasks and measures rather than whether hierarchical organisation was present in only one group.

The present data do not determine why the two age groups showed different patterns across the tasks. Because the study concerns the representation of algebraic syntax in memory and production, these differences should not be reduced to general developmental changes in executive control. One possibility is that age and experience affect how reliably operator-defined chunks are maintained and used during copying; this possibility requires further investigation.

To contextualise children's performance in the experimental tasks, we also examined their scores on two independent classroom worksheets: a ten-item order-of-operations arithmetic worksheet and a 36-item multiplication-fact worksheet, both completed separately from the copying tasks. Children performed near ceiling on both: mean accuracy was 93.7% ($SD = 11.6$, range 60–100.0%) on the arithmetic worksheet ($n = 27$), and 96.3% of the maximum score ($SD = 9.9$ points on a 90-point scale, range 45–90) on the multiplication worksheet ($n = 19$).

These worksheet results indicate that the children were generally familiar with the arithmetic knowledge required to interpret operator precedence. Therefore, the children's pattern in the copying tasks is unlikely to be explained simply by lack of basic arithmetic knowledge.

4.5 Theoretical and Educational Implications

Although the present study focused on algebraic expressions, the findings are relevant to broader questions concerning hierarchical symbolic processing. Similar organisational principles have been proposed in language, music, and action planning, suggesting that mathematical notation may rely on domain-general mechanisms for structuring complex symbolic sequences.

The findings also have potential implications for mathematics education. If hierarchical organisation naturally guides symbolic production, instructional approaches that explicitly emphasise meaningful structural groupings may help reduce planning demands during expression production. Because instructional methods were not evaluated directly, these implications should be viewed as hypotheses for future educational research rather than conclusions of the present study.

4.6 Open Questions

Several questions remain open. Some concern the interpretation of the present findings, whereas others point to methodological directions that could strengthen future work.

First, this study relied entirely on behavioural measures to infer internal representations. Behavioural data constrain but cannot directly reveal the underlying cognitive or neural processes; combining these paradigms with eye-tracking or electrophysiological measures would help distinguish between the representational and strategic accounts discussed above.

Second, the different accuracy patterns observed in children and adults across the two tasks underscore the need for a more comprehensive developmental trajectory than the two age groups tested here allow. A design spanning additional ages would help determine how structure-sensitive representation and production change with development.

Third, the two tasks isolate different parts of the pathway from encoding to production only imperfectly. The Delayed Copying Task places greater emphasis on maintaining the complete expression

in memory, whereas the Hidden Copying Task allows repeated visual access and therefore provides more opportunity to examine copying strategy. Future work could manipulate these stages more directly.

Fourth, the present expressions were deliberately simple, to keep operator structure isolated from other sources of complexity. Whether the same hierarchical organisation extends to expressions with more operators, nested structure, or genuine calculation rather than reproduction, remains to be tested.

4.7 Conclusion

This study examined whether hierarchical operator structure influences the representation and production of algebraic expressions, and whether these processes differ between children and adults. The results showed sensitivity to operator structure in both tasks, but in different measures and age groups: children's accuracy differed between structures in the Delayed Copying Task, whereas adults showed a structure-dependent accuracy difference in the Hidden Copying Task. The Hidden timing-error correlations also differed descriptively between the two critical conditions, but their direct difference was not statistically significant. Timing profiles showed descriptive trends consistent with the predicted structural organisation, although the critical interactions were not statistically significant. Together, these findings support the conclusion that operator structure can influence both the representation maintained in memory and the organisation of written production.

References

- Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 4, 417–423.
- Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–89). Academic Press.
- Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child Development*, 81, 1641–1660.
- Braithwaite, D. W., Goldstone, R. L., van der Maas, H. L. J., & Landy, D. H. (2016). Non-formal mechanisms in mathematical cognitive development: The case of arithmetic. *Cognition*, 149, 40–55.
- Chekaf, M., Cowan, N., & Mathy, F. (2016). Chunk formation in immediate memory and how it relates to data compression. *Cognition*, 155, 96–107. <https://doi.org/10.1016/j.cognition.2016.05.024>
- Chomsky, N. (1957). *Syntactic structures*. Mouton.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. MIT Press.
- Cowan, N. (2014). Working memory underpins cognitive development, learning, and education. *Educational Psychology Review*, 26, 197–223.
- Dotan, D., Diskin, R., Levy, R., & Yachini, M. (2020). WriTracker: Software for experimental investigation of handwriting. Tel Aviv University. <http://www.mathinklab.org/writracker>
- Fitch, W. T., & Martins, M. J. D. (2014). Hierarchical processing in music, language, and action: Lashley revisited. *Annals of the New York Academy of Sciences*, 1316, 87–104.
- Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, 40, 177–190.
- Gobet, F., Lane, P. C. R., Croker, S., Cheng, P. C.-H., Jones, G., Oliver, I., & Pine, J. M. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences*, 5, 236–243.
- Harrison, T. B., & Stiles, J. (2009). Hierarchical forms processing in adults and children. *Journal of Experimental Child Psychology*, 103, 222–240.
- Johnson-Laird, P. N. (1983). *Mental models*. Cambridge University Press.
- Kirshner, D. (1989). The visual syntax of algebra. *Journal for Research in Mathematics Education*, 20, 274–287.
- Landy, D., & Goldstone, R. L. (2007). How abstract is symbolic thought? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 720–733.
- Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112–136). Wiley.
- Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. MIT Press.

- Marghetis, T., Landy, D., & Goldstone, R. L. (2016). Mastering algebra retrains the visual system to perceive hierarchical structure in equations. *Cognitive Research: Principles and Implications*, 1(1), 25.
- Martins, M. J. D. (2025). From fractal geometry to fractal cognition: Experimental tools and future directions for studying recursive hierarchical embedding. *Fractal and Fractional*, 9(10), 654.
- Martins, M. J. D., Laaha, S., Freiberger, E. M., Choi, S., & Fitch, W. T. (2014). How children perceive fractals: Hierarchical self-similarity and cognitive development. *Cognition*, 133(1), 10–24.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81–97.
- Musfeld, P., Dutli, J., Oberauer, K., & Bartsch, L. M. (2024). Grouping in working memory guides chunk formation in long-term memory: Evidence from the Hebb effect. *Cognition*, 248, 105795.
- Palmer, S. E. (1999). *Vision science: Photons to phenomenology*. MIT Press.
- Rosenbaum, D. A., Cohen, R. G., Jax, S. A., Weiss, D. J., & van der Wel, R. (2007). The problem of serial order in behavior: Lashley's legacy. *Human Movement Science*, 26, 525–554.
- Scheepers, C., Sturt, P., Martin, C. J., Myachykov, A., Teevan, K., & Viskupova, I. (2011). Structural priming across cognitive domains: From simple arithmetic to relative-clause attachment. *Psychological Science*, 22(10), 1319–1326.
- Van de Cavey, J., & Hartsuiker, R. J. (2016). Is there a domain-general cognitive structuring system? Evidence from structural priming across music, math, action descriptions, and language. *Cognition*, 146, 172–184.

Appendix A. Z-Scored Inter-Token Gap Analyses

Z-scoring the inter-token gaps within participant reduces between-participant differences in overall timing scale. The z-scored analyses therefore provide a supplementary robustness check on the raw-time analyses reported in the main Results section. The critical Position $\times$ Condition effects remained non-significant after z-scoring (Delayed: children, $F(1, 22) = 2.75$, $p = .11$; adults, $F(1, 24) = 0.11$, $p = .75$; Hidden: children, $F(1, 13) = 1.21$, $p = .29$; adults, $F(1, 20) = 0.43$, $p = .52$).

Supplementary Statistical Tables

Table A1

Correlation Statistics Reported in the Hidden Copying Task

Task	Analysis	Group	Statistic	p	Source
Hidden Copying	Mean pre-production delay $\times$ error rate (1-2-1 condition)	Adults	$r = .54$	$p = .007$	Results; Fig. 2
Hidden Copying	Mean pre-production delay $\times$ error rate (2-2 condition)	Adults	$r = -.13$	$p = .54$	Results; Fig. 2
Hidden Copying	Difference between the two dependent correlations	Adults	One-tailed paired condition-label permutation test	$p = .06$	Results; Fig. 2

Note. The direct comparison of correlations is the one-tailed paired condition-label permutation test reported in the final Results.

Table A2

Z-Scored Position $\times$ Condition Robustness Tests

Task	Group	Comparison	Position $\times$ Condition	p	Sig.
Delayed Copying	Children	2-2 vs. 1-2-1	$F(1, 22) = 2.75$	$p = .11$	ns
Delayed Copying	Adults	2-2 vs. 1-2-1	$F(1, 24) = 0.11$	$p = .75$	ns
Hidden Copying	Children	2-2 vs. 1-2-1	$F(1, 13) = 1.21$	$p = .29$	ns
Hidden Copying	Adults	2-2 vs. 1-2-1	$F(1, 20) = 0.43$	$p = .52$	ns

Note. All four tests are the supplementary within-participant z-scored versions of the focused raw-gap Position $\times$ Condition analyses reported in the Results.

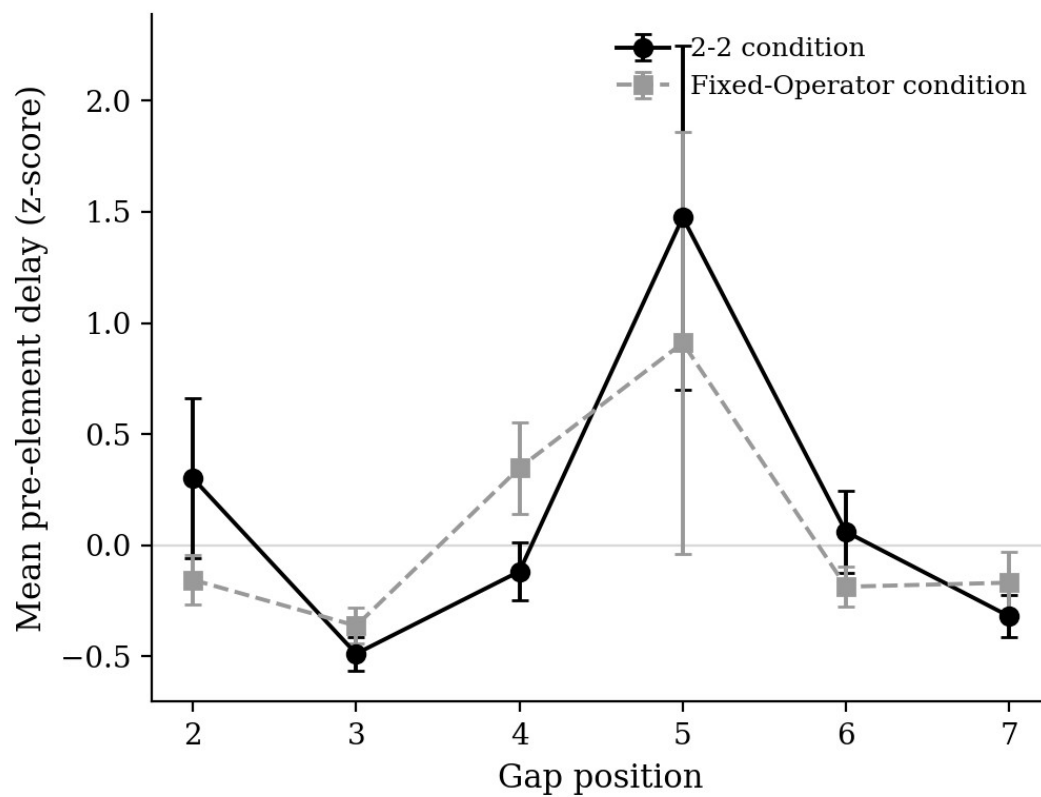


Figure A1. Delayed Copying Task, children. Z-scored inter-token-gap profile retained as the supplementary normalization analysis.

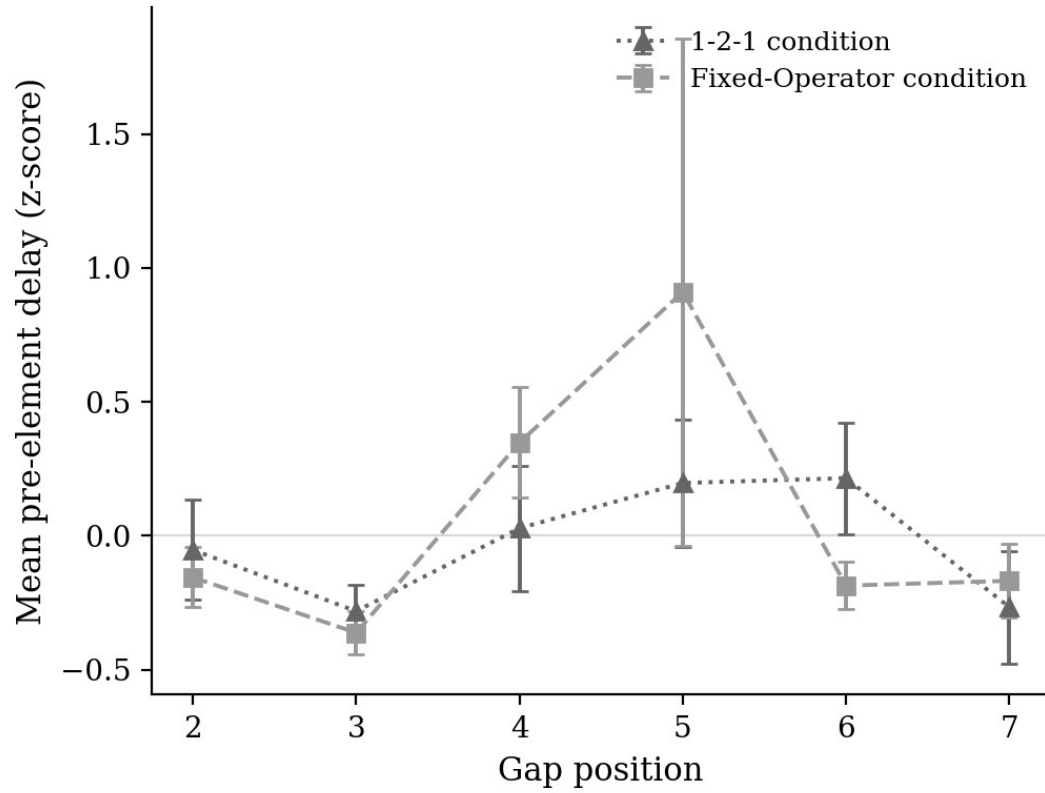


Figure A2. Delayed Copying Task, children, 1-2-1 vs. Fixed-Operator. Z-scored inter-token-gap profile retained as the supplementary normalization analysis.

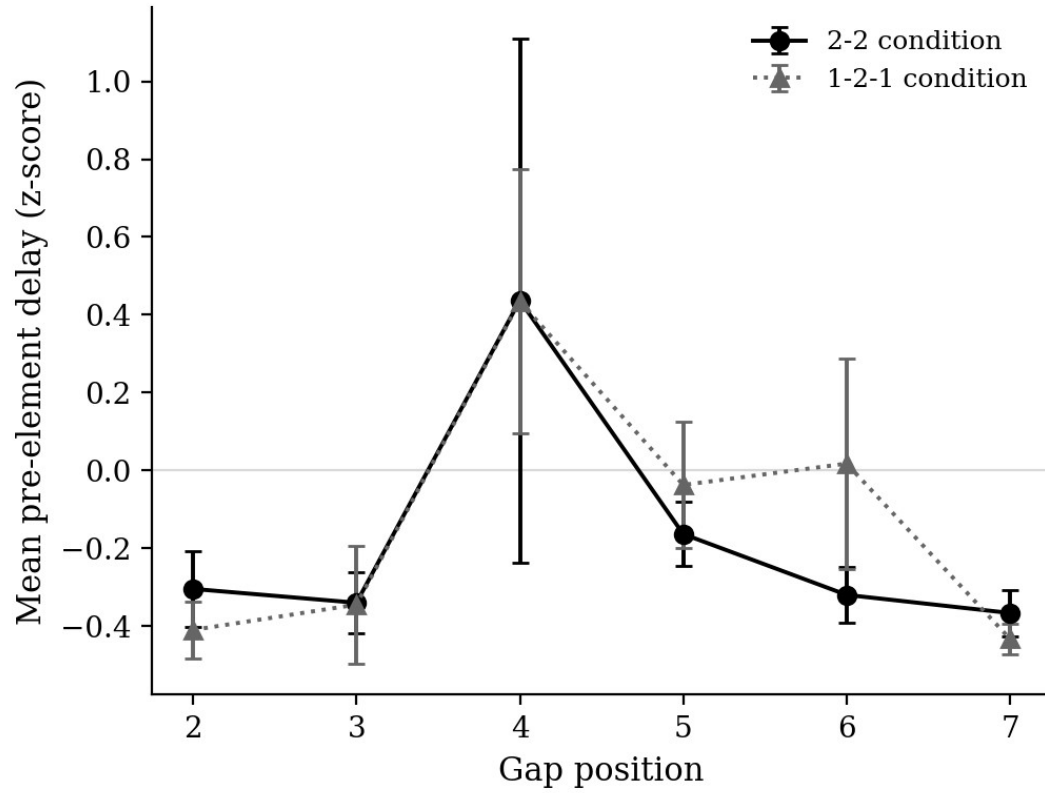


Figure A3. Delayed Copying Task, adults, 2-2 vs. 1-2-1. Z-scored inter-token-gap profile retained as the supplementary normalization analysis.

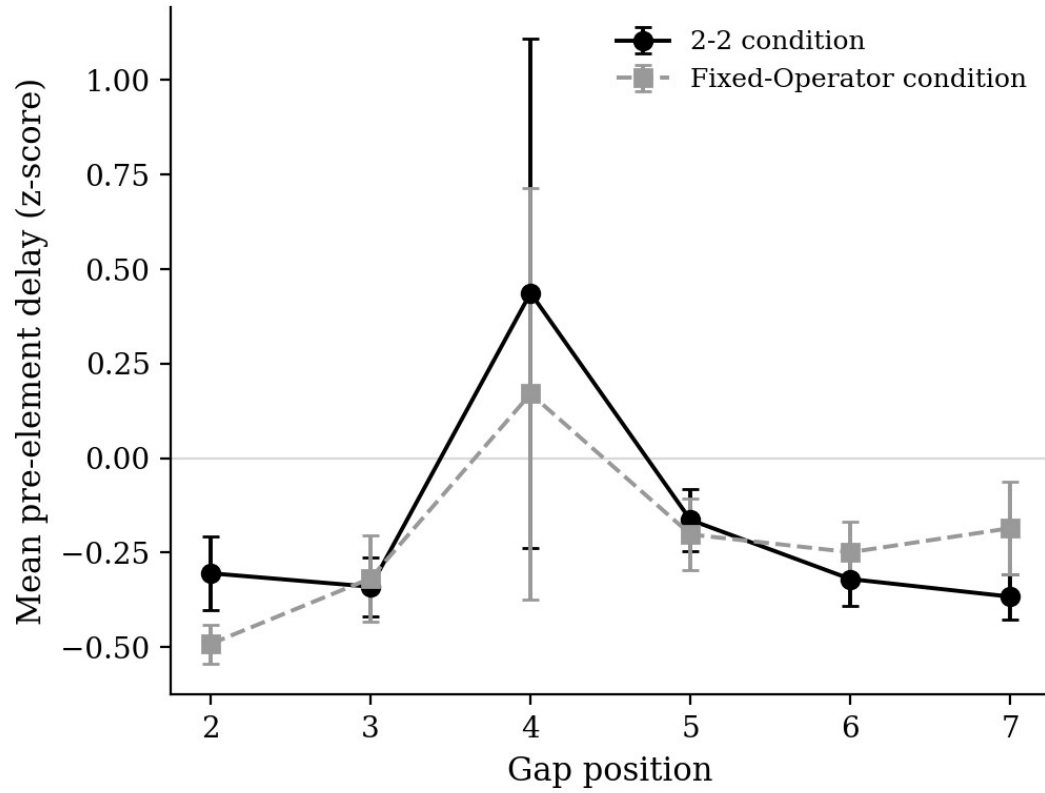


Figure A4. Delayed Copying Task, adults, 2-2 vs. Fixed-Operator. Z-scored inter-token-gap profile retained as the supplementary normalization analysis.

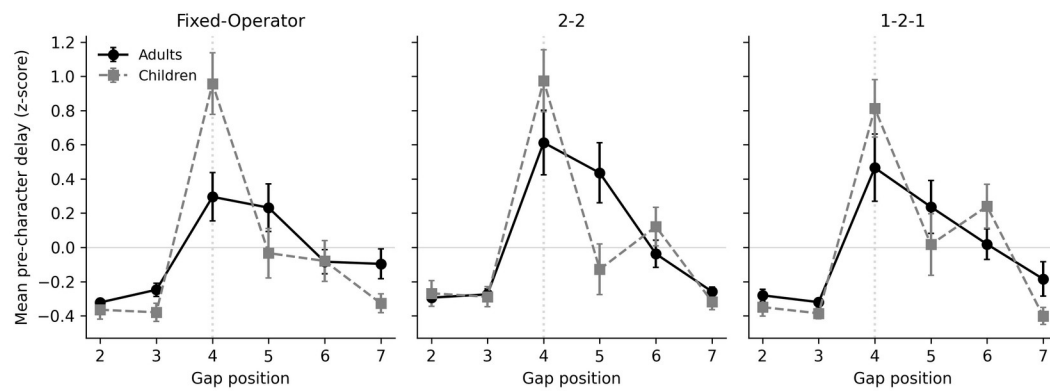


Figure A5. Hidden Copying Task. Z-scored inter-token-gap profile retained as the supplementary normalization analysis.