university of groningen

faculty of science and engineering

artificial intelligence

Exploring the Acquisition of Multiplication Facts in Primary Education: A Graph-Based Analysis of Skill Dependencies

Bachelor's Project Thesis

Ark Deliev, s4702328, a.deliev@student.rug.nl

Supervisor: Dr. T.M. (Teddy) Haile

Abstract: This research investigates the cognitive processes and skill acquisition mechanisms underlying multiplication fact learning in primary education through computational modelling techniques. By analysing performance data from 540 children (ages 6-10) who completed 315,690 trials using an adaptive learning system, we examined how young learners acquire multiplication facts and what skill dependency analysis reveals about their strategic approaches. The study employed TafelTrainer, an adaptive learning application implementing a three-level progression from initial calculation to automatised recall, combined with GraafTel, a novel graph-based algorithm for analysing skill hierarchies and dependencies. Through systematic analysis of response accuracy, reaction times, and memory decay patterns across learning stages, we identified three distinct skills with differential relationships to performance metrics. One skill maintained strong correlations with reaction times across all learning encounters, suggesting its association with calculation-based strategies that remain time-intensive despite practice. The other two skills demonstrated variable and generally decreasing correlations with reaction times in later encounters, indicating potential transition to retrieval-based processes that become more efficient with practice. These computational patterns align with established frameworks of cognitive skill acquisition that propose discrete transitions between computational, retrieval-based, and automatic processing during learning. The findings contribute to our understanding of mathematical cognition while informing the development of adaptive learning technologies that more precisely align with learners' cognitive developmental trajectories.

1 Introduction

Mastering fundamental arithmetic operations, such as multiplication, constitutes a necessary foundation for mathematical competence and the development of more advanced mathematical skills. Research by Kaskens et al. (2022) demonstrates that *arithmetic fluency*—specifically the ability to quickly and accurately retrieve arithmetic facts—plays a significant role in predicting mathematical problem-solving performance, particularly in solving multistep problems. These findings are consistent with the multi-level framework of mathematical cognition described by Gilmore (2023), which emphasises that arithmetic fluency is one of the key specific components of mathematics that contributes to overall mathematics achievement.

Significant challenges exist in developing arithmetic fluency, particularly in implementing effective instructional strategies that accommodate diverse cognitive profiles. Research by Das and Janzen (2004) identifies working memory and processing speed as key cognitive factors influencing mathematical learning, with these constraints affecting both computational efficiency and fact retrieval. These individual differences in cognitive processing necessitate adaptive learning approaches capable of responding to varied learning trajectories and strategy development patterns.

Traditional instructional methods, often relying on rote memorisation, may not optimally align with the cognitive processing of numerical information because they do not leverage the way humans naturally structure and retrieve mathematical knowledge. Research in numerical cognition suggests that mathematical learning involves both declarative memory (fact retrieval) and procedural strategies (rule-based problem-solving), with efficient problem-solving requiring flexible integration of these components (Dehaene, 2011). Rote memorisation primarily strengthens declarative recall but can bypass the conceptual understanding necessary for flexible application, limiting transferability to novel problems. Furthermore, studies indicate that numerical processing relies on the mental number line—a cognitive representation where numbers are spatially organised—and that meaningful engagement with numerical relationships, such as through strategy-based learning, better supports long-term retention and adaptive skill use (Siegler, 2009). Recent research indicates that approaches prioritising conceptual understanding and strategic problem-solving promote deeper learning, as they align with cognitive mechanisms underpinning numerical reasoning and retrieval (del Carmen Chamorro, 2021). By developing and implementing computational frameworks for tracking skill acquisition and strategy development, we can create more responsive educational tools that adapt to individual learning patterns while maintaining rigorous assessment of mastery progression.

This research examines the acquisition of multiplication facts through *computational modelling* of *learning trajectories* among primary school students aged 6-10. Computational modelling refers to the use of algorithmic and mathematical representations to simulate and analyse cognitive processes underlying learning. In this context, it enables the systematic study of how multiplication facts are acquired by constructing models that infer skill development patterns from student performance data. Learning trajectories describe the progressive paths through which students acquire mathematical knowledge, encompassing the sequence of concepts, strategies, and skills they develop over time. By utilising a large-scale dataset of 315,690 trials collected from 540 students through an adaptive digital learning environment that automatically adjusts practice schedules based on individual performance patterns, we apply graph-based analysis techniques to construct and analyse hierarchical skill structures, revealing patterns in fact mastery and strategy development. Our investigation addresses the following research question:

> *How do young learners in primary school (ages 6-10) acquire multiplication facts, and what can skill dependency analysis reveal about the strategies they use in mastering these facts?*

1.1 Knowledge Tracing and Adaptive Learning Systems

Knowledge tracing (KT) systems enable systematic monitoring of student learning progression through interactions with digital learning materials (Abdelrahman et al., 2023). These systems employ computational methods to observe, represent, and quantify students' knowledge states, often focusing on their mastery of specific skills underlying the educational content. Recent advances in KT have facilitated more precise modelling of learning patterns by accounting for complex factors such as skill dependencies and temporal dynamics in knowledge acquisition (van Rijn et al., 2009).

A key exemplar of these principles is TafelTrainer (Iancu et al., 2024; van der Velde, 2024), an adaptive learning system designed specifically for multiplication fact automatisation. TafelTrainer implements a three-tiered learning structure that guides students from initial exposure to rapid, accurate recall (Iancu et al., 2024). The system integrates cognitive modelling principles to track memory retention and forgetting, aligning practice schedules with the cognitive processes involved in learning. Rather than maintaining fixed intervals, TafelTrainer employs an adaptive spaced repetition strategy that continuously updates each learner's "speed of forgetting" parameter (α). This adaptive strategy ensures that practice occurs briefly before predicted memory decay (Iancu et al., 2024; van Rijn et al., 2009).

TafelTrainer's adaptive scheduling methodology is based on the latency-based adaptation model from van Rijn et al. (2009), which uses response latencies to estimate individualised forgetting rates for each multiplication fact. By analysing how quickly a learner responds to a retrieval attempt, the system infers the strength of the memory trace: faster responses indicate stronger memory, leading to a lower estimated forgetting rate and thus longer intervals before the next practice, while slower responses suggest weaker memory, resulting in a higher forgetting rate and more frequent practice. For example, if a student quickly recalls that $3 \times 4 = 12$, the system may extend the time before retesting this fact, whereas a delayed response for 7×8 might prompt earlier review. This dynamic adjustment ensures that practice occurs just before the predicted point of forgetting, optimising both the spacing and testing effects. Supporting this approach, Sense et al. (2016) demonstrated that forgetting rates can vary significantly across different materials for the same individual, underscoring the importance of fact-specific adaptations to accommodate varying cognitive demands.

Maths Garden, developed by Klinkenberg et al. (2011), complements these adaptive approaches by introducing a distinct methodology for tracking knowledge states through performance-driven parameter updates. While TafelTrainer focuses on memory retention through latency-based adaptation, Maths Garden employs a modified Elo rating system (Elo, 1978), originally designed for chess, to dynamically estimate learner ability and item difficulty. After each interaction, the system adjusts ratings based on discrepancies between expected and observed performance, incorporating response times to reward accurate answers given at developmentally appropriate speeds while penalising rapid errors. This approach operationalises knowledge tracing as a continuous calibration process, where iterative updates approximate learners' evolving mastery states. However, Maths Garden's linear rating model, though effective for adaptive practice scheduling, treats arithmetic facts as isolated units and may not fully capture the multidimensional nature of skill acquisition—particularly the interdependencies between facts and strategic flexibility emphasised by Siegler (2009).

GraafTel addresses this limitation by implementing data-driven graph-based skill structures (Taatgen & Blankestijn, 2024). Unlike systems like Maths Garden, which rely on linear rating models to rank items on a single difficulty dimension, GraafTel organises skills hierarchically in a graph where nodes represent clusters of items requiring specific skill combinations and edges denote prerequisite relationships inferred from student performance data. For instance, foundational facts like 2×3 might emerge as prerequisite nodes for solving derived problems like 4×3 if learners systematically apply doubling strategies. This methodology avoids rigid predefined skill taxonomies, instead inferring structures from empirical patterns in learner performance, thereby capturing multidimensional interdependencies—such as how known facts might scaffold new strategies—that linear models oversimplify. While GraafTel's inferred structures—such as clusters of interconnected skills, prerequisite dependencies between facts, or hierarchical skill progressions—can suggest how students develop and deploy mathematical strategies, interpreting these findings requires grounding in theoretical and empirical research on mathematical skill acquisition. The following section explores mathematical skill development and provides the context for enabling meaningful analysis of the strategic patterns that may emerge in GraafTel's skill graphs—for example, clusters indicating a transition from counting-based strategies (e.g., repeated addition for 3×4) to derived-fact usage (e.g., leveraging 2×6 to solve 4×6).

1.2 Learning of Basic Skills in Mathematics

The computational patterns revealed through systems like GraafTel must be interpreted within the broader context of cognitive development and strategic learning progressions in mathematics. Siegler (1988) established that children's multiplication skill development involves a sophisticated model of strategy choice rather than simple fact memorisation. Through systematic investigation of children's problem-solving approaches, Siegler (1988) identified multiple concurrent strategies that learners employ when solving multiplication problems.

Building on this foundation, Lemaire and Siegler (1995) conducted a longitudinal investigation of French second-graders' acquisition of single-digit multiplication skills, identifying four distinct dimensions of strategic change that contribute to improvements in speed and accuracy: (1) introduction of new strategies, (2) increasing use of more efficient existing strategies, (3) more efficient execution of each strategy, and (4) more adaptive choices among strategies. Their research demonstrated that even early in learning, children's strategy choices were highly adaptive, with retrieval dominating for easier problems while backup strategies like repeated addition were employed for more challenging ones. Importantly, Lemaire and Siegler (1995) found that children continued to use multiple strategies throughout the learning process, with the transition to retrieval-based approaches occurring gradually and selectively, based on problem characteristics rather than as a wholesale shift.

The process model proposed by Siegler (1988) describes three key mechanisms involved in multiplication problem-solving. First, learners may retrieve answers directly from memory based on the associative strength—the strength of the mental connection linking a multiplication problem (e.g., 3×4) to its correct answer (e.g., 12) in long-term memory—between problems and potential answers. When retrieval fails to meet a confidence criterion, learners employ backup strategies such as repeated addition or counting sets of objects. The choice between retrieval and backup strategies depends on factors including problem difficulty, the current associative strength of the problem-answer link (reflecting prior exposure and practice frequency), and previous experience (Siegler, 1988).

This framework helps explain observed patterns in children's multiplication performance. As noted by Siegler (1988), backup strategies are used more frequently on problems that elicit higher error rates and longer solution times. The model accounts for systematic variations in strategy use across different problem types—for instance, problems with larger operands tend to elicit more backup strategy use compared to problems with smaller operands. These patterns emerge from the interaction between associative strength, confidence criteria, and the relative success of different approaches.

Recent work by Gilmore (2023) builds on this understanding by examining how domain-specific skills like fact retrieval interact with domain-general cognitive processes such as working memory and inhibition. The constraints imposed by these cognitive factors, as identified by Das and Janzen (2004), influence both initial learning and subsequent strategy selection. Additionally, Campbell and Thompson (2012) highlights how strengthening certain fact associations may temporarily inhibit access to related facts through interference effects.

Understanding how learners navigate between retrieval attempts and backup strategies helps explain the clusters and relationships manifested in GraafTel's skill graphs. This perspective on strategy development naturally extends to broader theories of cognitive skill acquisition and the mechanisms underlying the transition from declarative to procedural knowledge.

1.3 Human Fact Learning

Siegler (1988) emphasises adaptive strategy use in mathematics learning, which broader frameworks extend to explain how components of a skill evolve into smooth, automatic performance. Anderson (1982) outlines a progression from declarative knowledge—where learners deliberately recall facts and rules—to procedural knowledge—where these elements are applied rapidly and with minimal cognitive load. In the early stages of learning multiplication, students might carefully break problems into smaller segments, effectively applying "declarative" strategies. For example, when confronted with 7×8, a novice learner might first calculate $7 \times 5 = 35$, then $7 \times 3 = 21$, and finally combine these results to arrive at 56. As practice continues, these procedures become integrated and streamlined, allowing instantaneous retrieval of facts and effortless application in problem-solving. This progression clarifies how learners transition from laborious, step-by-step reasoning to more direct and fluent recall.

Tenison et al. (2016) provide empirical support for this multiphase progression through their fMRI investigation of skill acquisition. Their research identified three distinct learning phases that align closely with the theoretical framework proposed by Anderson (1982): a computation-dominant phase, a retrieval-based phase, and an automatic response phase. By combining multi-voxel pattern analysis with hidden semi-Markov modelling, they demonstrated that the majority of practice-related speed up stems from discrete changes in cognitive processing—specifically, transitions between these learning phases—rather than from continuous improvements within each phase. Particularly noteworthy is their finding that the solving stage of the first learning phase involves a sequence of arithmetic computations, while the transition to the second phase occurs when learners can retrieve answers directly, substantially reducing solution time. With continued practice, learners then transition to the third phase where problems are recognised as unified patterns that trigger automatic responses. This neuroimaging evidence provides strong empirical validation for the theoretical phase transitions described in cognitive skill acquisition models.

Furthermore, Anderson (1982) describes processes of knowledge generalisation and discrimination that demonstrate how learners refine their multiplication skills over repeated encounters. Generalisation allows them to apply learned strategies to novel problems, while discrimination helps them recognise when certain approaches are inappropriate. For instance, a student who has mastered the strategy of doubling to calculate 4×7 (by computing 2×7 and then doubling the result) might generalise this approach when encountering 4×9. Conversely, discrimination occurs when the student recognises that using the doubling strategy for 7×8 is less efficient than direct retrieval after sufficient practice. This iterative process of testing, adjusting, and internalising strategies ensures that learners become adept at selecting the best tools for each scenario. Anderson's framework provides a temporal perspective on skill acquisition, explaining changes in performance and strategy use across learning encounters and difficulty levels.

Separate from Anderson's framework, Taatgen (2013) introduces the primitive elements theory (PRIMs), suggesting that cognitive skills can be decomposed into fundamental processing units. These primitive elements serve as the lowest level building blocks from which more complex cognitive operations are constructed. According to this theory, PRIMs—such as recognising numerals or performing basic comparisons—constitute essential cognitive operations that, when combined, enable the execution of more sophisticated mathematical procedures. For example, the cognitive operation of recognising that 5×8 equals 40 involves primitive elements such as visual processing of numerals, retrieval of arithmetic facts, and selection of the appropriate response. The PRIMs framework provides a mechanistic explanation for how basic cognitive operations can be reused across different tasks, facilitating transfer of learning.

While PRIMs operate at the most fundamental level of cognitive processing, Taatgen and Blankestijn (2024) demonstrate that mathematical skill acquisition can be more effectively understood through higher-level skill structures that emerge from these primitive operations. Their data-driven approach to cognitive skill modelling reveals that mathematical competencies manifest as combinational and transferable skills that form hierarchical relationships. Taatgen and Blankestijn (2024) identify specific mathematical skills—such as simple arithmetic operations, handling large numbers, solving story problems, and multistep reasoning—that function as distinct, reusable components within a skill hierarchy. For instance, the skill of multiplying by 10 (e.g., $10 \times 6 = 60$) can be combined with addition to solve more complex problems like 12×6 by computing $(10 \times 6) + (2 \times 6) = 60 + 12 = 72$. Unlike PRIMs, which represent the most basic cognitive operations, these mathematical skills represent clusters of coordinated processes that learners combine to address increasingly complex problems.

This hierarchical skill perspective helps explain why certain clusters of related multiplication facts emerge in learning trajectories. For instance, mastery of basic multiplication facts (e.g., 2×3) functions as a foundational skill that can be combined with other skills (like doubling) to derive solutions to more complex problems (e.g., 4×3). A concrete manifestation of this hierarchy appears when students leverage knowledge of multiplication by 5 to compute products involving 6: having memorised that $5 \times 7 = 35$, a student might calculate 6×7 as $35 + 7 = 42$. Understanding the higher-level organisation of mathematical skills provides insight into how learners navigate their developing skill network, utilising existing components at various levels of abstraction to develop more sophisticated problem-solving approaches.

In essence, Anderson (1982) skill acquisition framework offers a temporal perspective on how learners progress from declarative to procedural knowledge across learning encounters, while Taatgen's models—from primitive elements to hierarchical skill structures—provide a structural account of how cognitive skills are organised and combined. Together, these complementary theoretical perspectives establish a methodological basis for examining both temporal progression in skill acquisition and the interconnected relationships between skills, thereby informing our empirical investigation. The analytical tools employed in this study, namely GraafTel, derive methodological utility from both frameworks: temporal changes across encounters are interpreted through Anderson's lens, while skill hierarchies and transfer relationships align with Taatgen's approach.

1.4 Outline of the Study

This research utilises comprehensive student performance data from an adaptive learning system designed to develop multiplication fact fluency among primary school students. Through the analysis of accuracy metrics, response times, and forgetting rates, this study implements a GraafTel-based skill analysis to investigate patterns in strategy development and fact mastery acquisition.

Building on previous research in cognitive skill acquisition (Anderson, 1982; Siegler, 1988), we identify three primary strategies that learners may employ when solving multiplication problems:

- *Counting-based strategies*, where early learners prioritise accuracy over speed by using repeated addition (e.g., solving 3 × 4 as 4 + 4 + 4) (Siegler, 1988)
- *Derived-facts strategies*, which align with Taatgen's primitive elements theory (Taatgen, 2013), where learners reuse and adapt known facts to solve new problems (e.g., using 2 × 3 = 6 to deduce 4 × 3 = 12 by doubling)
- *Direct fact retrieval*, representing the transition from declarative knowledge to procedural fluency described by Anderson (1982), where answers are automatically retrieved without calculation

Our central research question remains: *How do young learners in primary school (ages 6-10) acquire multiplication facts, and what can skill dependency analysis reveal about the strategies they use in mastering these facts?* To address this question, we analyse performance data from 540 students across three distinct levels of the TafelTrainer platform, capturing 315,690 trials over 17,575 sessions. For each student-level combination, we examine three critical learning stages: initial encounter, middle encounter, and final encounter with multiplication facts. This temporal approach allows us to trace developmental trajectories and strategic shifts as learners progress from novice to more expert performance.

We propose three main hypotheses regarding skill acquisition patterns:

1. *Skill analysis can distinguish between retrieval and calculation strategies.* We test this by examining correlations between GraafTel skill probabilities and reaction times across different learning encounters. If distinct cognitive processes are captured by different skills, we expect differential correlation patterns—specifically, skills associated with calculation should maintain strong positive correlations with reaction time throughout learning, while retrieval-based skills should show decreasing correlations as automaticity develops.
2. *Skill analysis measures independent cognitive processes rather than redundant metrics.* We evaluate this by analysing inter-skill correlations across different skill configurations (3-6 skills). If the model successfully captures distinct cognitive processes, we expect relatively low inter-skill correlations, particularly in configurations that optimally balance model complexity and predictive accuracy.
3. *Skill analysis reveals the development of strategic approaches across learning.* We test this by comparing skill probability distributions across different levels and encounters, particularly focusing on how skill-reaction time relationships evolve. We expect to observe systematic shifts from calculation-dominant to retrieval-dominant patterns, evidenced by changing correlations between skill probabilities and performance metrics.

The unique contribution of this research lies in its application of a novel graph-based skill analysis methodology (GraafTel) to empirically derived learning trajectories, providing a data-driven approach to understanding multiplication fact acquisition that extends beyond traditional theoretical models. Unlike previous studies that often rely on predefined skill taxonomies or linear difficulty assessments, our approach infers skill structures directly from performance patterns, capturing the multidimensional nature of mathematical learning. By integrating temporal analysis (across learning encounters) with structural analysis (skill dependencies and hierarchies), we bridge theoretical frameworks of cognitive skill acquisition with empirical patterns in naturalistic learning environments.

This approach has significant educational implications, potentially informing the development of adaptive learning technologies that more precisely align with learners' cognitive development trajectories. Rather than treating multiplication facts as isolated units to be memorised, our methodology reveals the interconnected skill structures that underpin effective learning, enabling more targeted instructional interventions that leverage natural learning progressions and strategic development patterns.

Through this comprehensive analysis of multiplication fact acquisition, we aim to advance both theoretical understanding of mathematical cognition and practical applications in educational technology design, contributing to more effective and developmentally appropriate learning experiences for young mathematics students.

2 Methods

2.1 Data Collection

The dataset was sourced through Iancu et al. (2024), who conducted a naturalistic study using the TafelTrainer adaptive learning system across 11 primary schools in the Netherlands. The study involved 540 students aged 6–10 years, who collectively completed 315,690 trials across 17,575 sessions during a 197-day period. While teachers retained autonomy over implementation scheduling, students maintained freedom to select multiplication tables and difficulty levels according to their preferences (Iancu et al., 2024). Average platform engagement was 66 minutes per student, distributed across approximately 26 sessions.

2.2 Computational Framework

The study utilised two primary computational systems: TafelTrainer for adaptive learning delivery (Iancu et al., 2024) and GraafTel for skill dependency analysis.

2.2.1 TafelTrainer Implementation

As described by Iancu et al. (2024), TafelTrainer implemented a three-tiered learning progression. The first level, focused on foundation building, presented facts sequentially without algorithmic adaptation and required a single correct response for completion. The second level introduced adaptive practice, employing memory model-based scheduling using accuracy metrics and implementing model-based mastery criteria. The third level aimed at automatisation by adding an 8-second response limit while maintaining the adaptive scheduling and model-based mastery assessment from the previous level.

The system's memory model predicted fact activation levels through:

$$M(f,t) = A(f, t + 24h) \geq \tau_M \tag{2.1}$$

$$= \ln \sum_j ((t + 24h) - t_j)^{-d} \geq \tau_M \tag{2.2}$$

In this equation, $M(f,t)$ represents the predicted mastery of fact f at time t, while $A(f, t + 24h)$ indicates the activation level prediction 24 hours ahead. The parameter τ_M serves as the activation threshold, and d represents the decay parameter, which was fixed at 0.45.

The model incorporated fact-specific forgetting rates (α), which were updated based on student performance. In Level 2, α adjustments were based solely on response accuracy, with increments or decrements of 0.01 from the default assumption of 0.3. For Level 3, the α adjustments incorporated both accuracy and response time, allowing for more nuanced adaptation to student performance.

2.2.2 GraafTel Algorithm

The GraafTel system employs a novel graph-based algorithm, currently under development by Taatgen and colleagues, to analyse skill hierarchies and dependencies. This methodology, part of an ongoing research initiative to advance adaptive learning frameworks, represents both items and learners through skill vectors whose dimensionality corresponds to the number of skills in the domain. For items, vector elements represent the probability that a specific skill is required for successful completion, while for learners, these elements indicate the probability of skill mastery.

The system calculates the expected probability of success (ES) through the following multiplicative model:

$$ES = \prod_i (1 - x_i + x_i s_i) \tag{2.3}$$

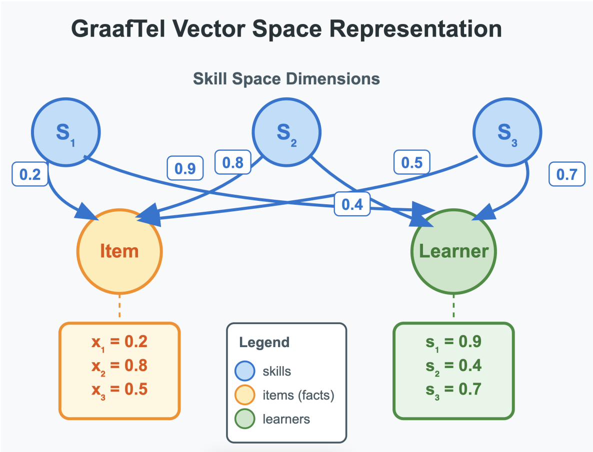

Figure 2.1: Illustration of the GraafTel vector representation model. Both multiplication facts and learners are represented through skill vectors. For items, vector elements (x_i) represent the probability that a specific skill is required for successful completion; for learners, elements (s_i) indicate the probability of skill mastery.

In this equation, x represents the item skill vector indicating the probability of skill requirements, while s denotes the learner skill vector reflecting the probability of skill mastery. Both vectors were constrained to values between 0 and 1.

The model evaluates performance by comparing actual scores with expected probabilities:

$$error = score - ES \tag{2.4}$$

This error is then used to update the skill vectors x and s via a gradient-based optimisation procedure. The updates adjust each vector component proportionally to its contribution to the expected success probability, aiming to minimise the discrepancy between predicted and actual performance. Specifically, the adjustment process leverages partial derivatives of ES with respect to x_i and s_i, scaled by the error and distinct learning rates (α_x and α_s) for items and students, respectively. To enhance efficiency and stability, the ADAM optimisation algorithm is employed, adapting the learning rates based on the first and second moments of the gradients (Kingma & Ba, 2017).

GraafTel Free Parameters The GraafTel algorithm requires setting several free parameters that influence the model's behaviour. The most critical parameter is the number of skills, which determines the dimensionality of the skill vectors and consequently the complexity of the skill structure that can be captured. We systematically evaluated skill configurations ranging from 3 to 6 skills to determine the optimal setting, assessing each configuration based on both model error rates and the independence of the resulting skill dimensions.

Additionally, the model's training process is governed by epoch configurations that determine how many iterations of the optimisation algorithm are performed. For this study, we experimented with epoch settings of 1000, 2000, 3000 and 4000, examining convergence patterns and stability of results. The final analysis employed 2000 epochs, which provided a suitable balance between computational efficiency and model convergence.

The present study contributes to a series of investigations testing GraafTel's capabilities in modelling strategic learning pathways.

2.3 Data Structure

For each student-level combination, we constructed three distinct datasets capturing different stages of learning. The initial encounter dataset contained the first attempt at each multiplication fact, while the middle encounter occurred half-way between the first and last trial per level, with adjustments to ensure distinction from the initial encounter. For instance, if a student did 6 trials, the 4th trial was taken as the middle encounter. The final encounter dataset comprised the most recent attempt at each fact. Response parameters included accuracy for all levels, with completion time measurements and memory decay rates only available for Levels 2 and 3.

GraafTel outputs skill dependency analyses with two entity types: multiplication facts (e.g., "1 × 1") and anonymised students. Each entry features a skill vector whose dimensionality matches the configuration (3–6 skills in our study). For items, vector elements represent the probability of requiring each skill for problem-solving (x_i in (2.3)); for students, elements denote inferred mastery probabilities (s_i). These bounded [0,1] values are iteratively refined through GraafTel's optimisation process, which minimises discrepancies between predicted and observed performance.

2.4 Data and Code Availability

To ensure research transparency and reproducibility, all data and analysis code used in this study have been made publicly available. The dataset containing anonymised student performance metrics, alongside the R scripts implementing the analyses presented in this paper, can be accessed through a public GitHub repository: https://github.com/theodrosmhaile/times_table_tracing. This repository includes the complete computational procedures for preprocessing the TafelTrainer performance data and generating the statistical results and visualisations presented in this study.

3 Results

Our analysis revealed distinct patterns in multiplication fact acquisition across learning stages and strategy development.

3.1 Overview of Learning Results

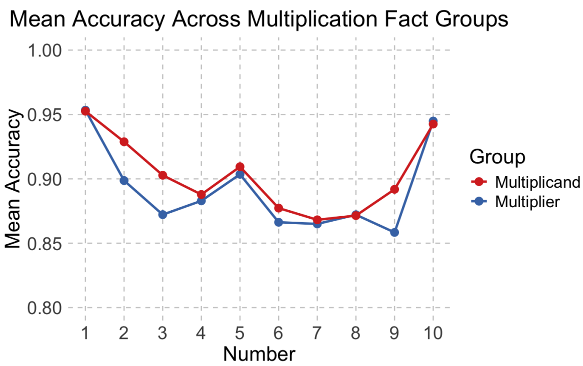

Figure 3.1: Mean accuracy across multiplication fact groups. The plot compares mean accuracy for multiplication facts grouped by multiplicand (red) and multiplier (blue). The x-axis represents the number in the multiplication fact (1–10), while the y-axis represents mean accuracy. Accuracy is highest for facts involving 1s and 10s, while mid-range numbers (3–7) show a dip in performance. The trend is similar for both multiplicand and multiplier groupings, though multiplicands generally show slightly higher accuracy.

Results vary systematically depending on columns of multiplication fact (e.g., column of numbers multiplied by 1, 2, 3 and so on) and the three conditions (L1, L2 and L3). As shown in Figure B.1 in Appendix B, accuracy is highest for multiplication facts involving multipliers 1 (Mean: 0.95, SD: 0.21), 2 (Mean: 0.90, SD: 0.30), 5 (Mean: 0.90, SD: 0.30), and 10 (Mean: 0.94, SD: 0.23). Figure 3.1 further illustrates these differences, showing mean accuracy for multiplicands (red) and multipliers (blue). While 1s and 10s consistently yield the highest accuracy, accuracy dips for mid-range numbers (particularly 3s to 7s), though the trend is not strictly linear. Notably, 5s show relatively high accuracy despite being in the mid-range group.

Reaction times varied considerably across levels and encounters, reflecting the influence of the adaptive learning system and time constraints. Overall, the mean reaction time across all trials was 5.16s (SD = 43.35s), with a median of 2.51s, indicating a positively skewed distribution. Notably, Level 2 exhibited substantially longer mean response times (M = 7.33s, SD = 60.11s, median = 2.80s) compared to Level 3 (M = 2.99s, SD = 11.83s, median = 2.29s). This difference likely reflects the 8-second time limit introduced in Level 3, which encouraged faster responses.

Across different encounter positions, reaction times showed modest variations, with first encounters averaging 5.33s (SD = 31.32s), middle encounters 4.69s (SD = 24.89s), and last encounters 5.45s (SD = 63.53s). The combination of level and encounter position revealed that Level 2 first encounters had the slowest responses (M = 7.60s, SD = 44.16s), while Level 3 last encounters showed the fastest performance (M = 3.02s, SD = 15.22s). Correct responses were substantially faster (M = 4.92s, SD = 42.14s, median = 2.44s) than incorrect ones (M = 7.23s, SD = 52.66s, median = 4.36s), suggesting that slower responses were often associated with difficulty or uncertainty.

Level	First	Middle	Last
1	0.88 ± 0.32	0.91 ± 0.28	0.95 ± 0.23
2	0.91 ± 0.28	0.92 ± 0.26	0.95 ± 0.21
3	0.84 ± 0.36	0.88 ± 0.33	0.87 ± 0.33

Table 3.1: Mean accuracy (± standard deviation) for each level and encounter combination.

Accuracy generally increases with later encounters across all levels, though with notable vari-

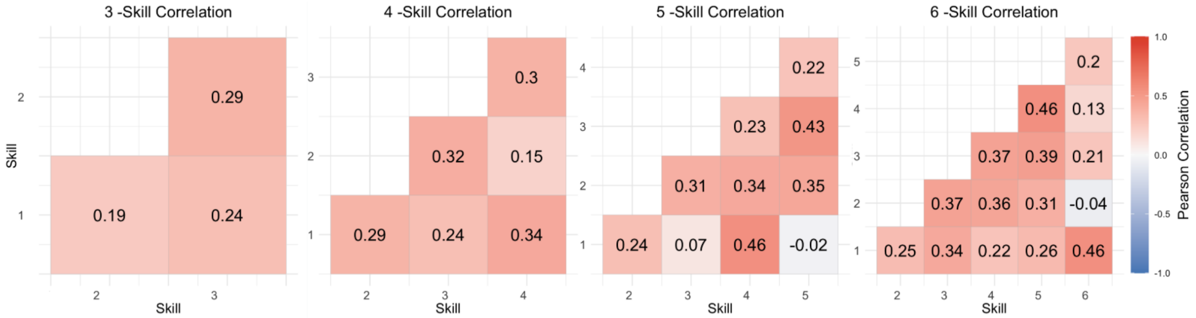

Figure 3.2: Correlation matrices of skills across different skill configurations. Each subplot represents the Pearson correlation coefficients between skills in configurations of 3, 4, 5, and 6 skills. The x-axis and y-axis represent skill indices, with each cell indicating the correlation between a pair of skills. The colour scale ranges from -1 (negative correlation, blue) to +1 (positive correlation, red), with white indicating no correlation. Lower correlation values suggest greater independence between measured skills, whereas higher values indicate stronger relationships between skill pairs.

ations. Level 1 demonstrates solid performance (mean = 91.30%, SD = 28.18%), while Level 2 shows the highest overall accuracy (mean = 93.00%, SD = 25.51%), likely attributable to the adaptive algorithm's effectiveness in optimising practice intervals. A slight decrease in accuracy occurs at Level 3 (mean = 86.46%, SD = 34.22%), which can be attributed to the introduction of the 8-second time limit constraint. It is important to note that all responses were included in these analyses, including those outside the expected 1-100 range. While excluding these potential 'skips' would have increased the baseline accuracy substantially (from approximately 61% to 70%), we determined that including all responses was methodologically appropriate, as there is no reliable method to differentiate between genuine calculation attempts, retrieval failures, or disengaged responses. This comprehensive approach provides a more complete assessment of performance patterns across the learning conditions.

3.2 Determining Optimal Skill Count for GraafTel Model

Analysis of skill correlations reveals that correlation strength varies systematically with the number of skills modelled in GraafTel. The three-skill configuration shows the lowest inter-skill correlations (r = 0.19 to 0.29), while configurations with more skills demonstrate progressively higher correlations. In the five- and six-skill configurations, some skill pairs show correlations as high as $r = 0.46$, suggesting increased overlap in measured mathematical strategies as the number of skills increases. This pattern supports the selection of the three-skill configuration for subsequent analyses, as it best maintains strategy independence while minimising average error rates during GraafTel model runs across various epoch configurations.

3.3 Correlation Between GraafTel Skill Ratings and TafelTrainer Reaction Time (RT)

To test our first hypothesis—that skill analysis can distinguish between retrieval and calculation strategies—we examined correlations between GraafTel skill probabilities and reaction times across different learning encounters. This analysis required integrating two primary data sources: (1) the GraafTel-generated item skill vectors (x_i) representing the probability that each skill is required for solving each multiplication fact, and (2) the mean reaction times for each multiplication fact at each encounter position within Levels 2 and 3. Level 1 data were excluded from this analysis as reaction times were not recorded in that level.

For each level-encounter combination, we calculated mean reaction times per multiplication fact and then computed Pearson correlation coefficients between each skill probability and these mean reaction times. To maintain statistical rigour, we applied Bonferroni correction to account for multiple comparisons across the three skills.

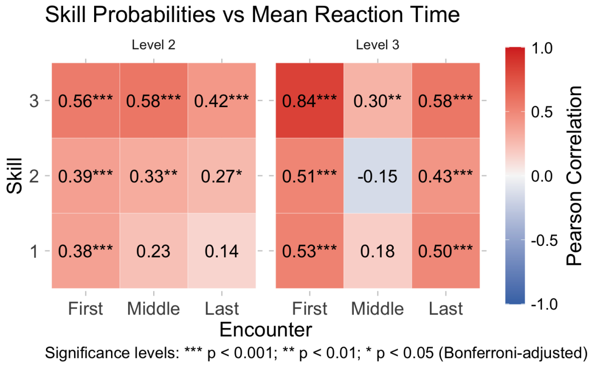

Figure 3.3: Pearson correlation coefficients between skill probabilities and mean reaction time (RT) across levels and encounters. The x-axis represents the encounter position (First, Middle, Last) within Level 2 and Level 3, while the y-axis represents different skills labelled by GraafTel. Each cell displays the correlation coefficient between the probability of a skill being required for an item and the corresponding mean RT at that encounter. Skill probabilities were derived from the *item* skill vector (x_i in Equation (2.3)) in the GraafTel model, representing the likelihood that a multiplication fact requires a given skill (see sections 2.2.2 and 2.3). The colour scale ranges from -1 (negative correlation, blue) to +1 (positive correlation, red), with white indicating no correlation. Significant correlations are marked with asterisks (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, Bonferroni-adjusted).

Figure 3.3 presents the correlation coefficients for each skill across all level-encounter combinations. Skill 3 shows generally strong correlations across most encounters and levels, ranging from $r = 0.30$ ($p < 0.01$) in the middle encounter of Level 3 to $r = 0.84$ ($p < 0.001$) in the first encounter of Level 3. Skills 1 and 2 show more variable correlation patterns. In Level 2, Skill 1's correlation decreases from a significant $r = 0.38$ ($p < 0.001$) in the first encounter to a non-significant $r = 0.14$ in the last encounter. Similarly, Skill 2's correlation weakens from $r = 0.39$ ($p < 0.001$) to $r = 0.27$ ($p < 0.05$). In Level 3, both Skills 1 and 2 show significant correlations in the first encounter ($r = 0.53$ and $r = 0.51$ respectively, both $p < 0.001$), non-significant correlations in the middle encounter, and significant correlations again in the last encounter ($r = 0.50$ and $r = 0.43$ respectively, both $p < 0.001$). Notably, Skill 2 shows a slight negative correlation ($r = -0.15$, non-significant) in the middle encounter of Level 3.

3.4 Correlation Between Level 1 Skill Probabilities and Level 3 Memory Decay Rates

The memory decay rate parameter (α) in TafelTrainer quantifies how quickly learned information fades from memory, with higher values indicating more rapid forgetting. Overall, the mean α value across all analysed data was 0.285 (SD = 0.047), with values ranging from 0.15 to 0.50, suggesting moderate memory stability across the student population.

Analysis by level revealed slightly different forgetting rates between Level 2 (M = 0.294, SD = 0.020) and Level 3 (M = 0.276, SD = 0.062), with the marginally lower mean in Level 3 potentially indicating improved retention with practice. Across encounter positions, middle encounters showed the highest average α (M = 0.292, SD = 0.042) compared to first (M = 0.284, SD = 0.044) and last encounters (M = 0.280, SD = 0.053). This pattern suggests a non-linear progression in memory decay rates during the learning process.

The combination of level and encounter position revealed that Level 2 first encounters showed the highest memory decay rate (M = 0.297, SD = 0.016), while Level 3 first encounters exhibited the lowest (M = 0.271, SD = 0.057). This contrast between levels may reflect different cognitive demands and strategy deployment between the accuracy-only focus of Level 2 versus the accuracy and speed requirements of Level 3.

To further investigate our first hypothesis—distinguishing between retrieval and calculation strategies—and to examine how early skill patterns might predict later memory characteristics, we analysed the relationship between skill probabilities from Level 1 and memory decay rates (α) from Level 3. This cross-level analysis provides insights into how the cognitive strategies employed during initial learning might relate to memory consolidation in later, more advanced learning stages.

Unlike the previous analysis that examined correlations within the same level, this analysis integrated data across different learning levels. We conducted two separate correlation analyses: (1) a fact-level analysis examining how skill requirements for specific multiplication facts relate to their memory characteristics, and (2) a student-level analysis investigating whether individual differences in skill profiles correlate with memory parameters.

For the fact-level analysis, we combined the GraafTel-generated item skill vectors (x_i) from Level 1, representing early strategy patterns, with the mean memory decay rates (α) for each multiplication fact at each encounter position in Level 3. The memory decay rate (α) quantifies how quickly learned information fades from memory, with higher values indicating more rapid forgetting.

For each encounter position in Level 3 (first, middle, last), we computed the mean α value for each multiplication fact and calculated Pearson correlation coefficients between these values and the corresponding Level 1 skill probabilities. To account for multiple comparisons, we applied Bonferroni correction with an adjustment factor of 3 (for the three skills tested simultaneously).

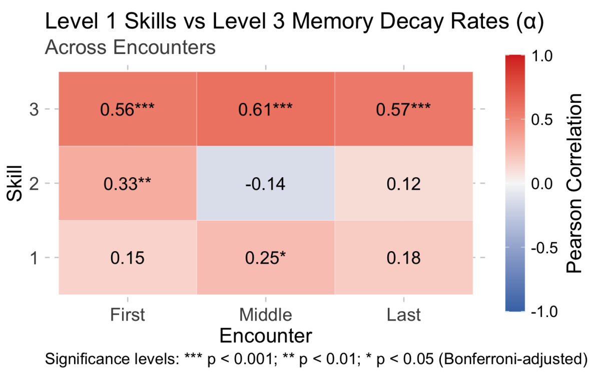

Figure 3.4: Pearson correlation coefficients between Level 1 skill probabilities and Level 3 memory decay rates (α) across different encounters. The x-axis represents the encounter position (First, Middle, Last) in Level 3, while the y-axis represents different skill levels from Level 1. Each cell displays the correlation coefficient between a specific Level 1 skill and the memory decay rate (α) at the corresponding encounter position in Level 3. The colour scale ranges from -1 (strong negative correlation, blue) to +1 (strong positive correlation, red), with white indicating no correlation. Significant correlations are marked with asterisks (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, Bonferroni-adjusted).

Figure 3.4 presents the correlation patterns between Level 1 skills and Level 3 memory decay rates at the fact level. Skill 3 demonstrates consistently moderate and highly significant positive correlations with memory decay rates across all encounters ($r = 0.56$, $r = 0.61$, and $r = 0.57$ for first, middle, and last encounters, respectively; all $p < 0.001$). This suggests that facts requiring higher Skill 3 probability in Level 1 tend to exhibit higher forgetting rates in Level 3. In contrast, Skills 1 and 2 show weaker and more variable correlations with memory decay rates. Skill 1 displays modest positive correlations across all encounters, with only the middle encounter reaching statistical significance ($r = 0.15$, non-significant; $r = 0.25$, $p < 0.05$; and $r = 0.18$, non-significant). Skill 2 shows a moderate significant positive correlation in the first encounter ($r = 0.33$, $p < 0.01$), a non-significant negative correlation in the middle encounter ($r = -0.14$), and a non-significant positive correlation in the last encounter ($r = 0.12$).

For the student-level analysis, we used student skill vectors (s_i) from Level 1 and their average memory decay rates across all facts in Level 3. This analysis revealed negligible correlations across all skills (r ranging from -0.09 to 0.04), with none reaching statistical significance. This contrast between fact-level and student-level analyses suggests either that skill ratings are independent of individual memory decay rates, that there are limitations in measuring student-level memory decay parameters, or that the relationship between skills and memory operates primarily at the item level rather than at the student level.

4 Discussion

4.1 Summary of Key Findings

This research aimed to understand how young learners in primary school (ages 6-10) acquire multiplication facts and to determine what skill dependency analysis could reveal about the strategies they use in mastering these facts. Through analysis of performance data from 540 students using the TafelTrainer adaptive learning system, we sought to identify patterns in strategy development and skill acquisition across different learning phases. By applying the GraafTel computational modelling approach to construct skill hierarchies, we examined how strategy use evolves from initial exposure to more advanced stages of learning.

Our first hypothesis proposed that skill analysis might distinguish between memory retrieval and calculation-based strategies. The correlation patterns between skill probabilities and response times support this differentiation, consistent with the strategy choice framework developed by Siegler (1988). The observation that Skill 3 maintains relatively consistent correlations with response times ($r = 0.56$ to 0.84) across encounters suggests this skill may correspond to calculation-based strategies that require greater cognitive resources and time. In contrast, the weakening correlations observed for Skills 1 and 2 in later encounters could indicate these skills relate to retrieval-based strategies that become more efficient with practice. These patterns align with Siegler (1988)'s finding that students employ multiple strategies when solving multiplication problems, with strategy selection being influenced by factors such as problem difficulty and previous experience.

The second hypothesis addressed the independence of measured skills. The three-skill configuration demonstrated relatively low inter-skill correlations ($r = 0.19$ to 0.29), suggesting that our chosen parameter value for the number of skills adequately captures distinct cognitive processes without redundancy. While this finding supports the methodological decision to use three skills in our analysis, it does not itself provide evidence about skill compilation or integration. To properly examine such cognitive processes, future research should track individual facts across encounters and learning levels, analysing whether the probabilities of specific skills (particularly those potentially associated with automatic retrieval) increase systematically with practice. In particular, if one skill represents fast automatic retrieval, we would expect to see its probability increase for multiplication facts as students progress from Level 1 first encounters to Level 3 final encounters.

Our third hypothesis concerned strategy development patterns. The differential correlation patterns between Skill 3 and Skills 1/2 with response times suggest potential differences in cognitive processing across learning encounters. Specifically, our results showed that Skill 3 maintained strong correlations with response times across all encounters ($r = 0.56$ to 0.84), while Skills 1 and 2 exhibited more variable and generally decreasing correlations in later encounters, particularly in Level 2. These empirical patterns align with theoretical frameworks of strategy development, such as those described by Siegler (1988), though our current analysis cannot definitively identify the specific cognitive mechanisms underlying these patterns. The observed correlation patterns provide preliminary evidence for differentiation between skills that remain consistently associated with longer response times (potentially calculation-based processes) and skills that become less associated with response times as learning progresses (potentially retrieval-based processes). To strengthen these interpretations, future research should track how individual facts' skill probability distributions evolve across learning encounters and directly correlate these changes with observed strategy use, perhaps through concurrent verbal protocols or other direct measures of strategy selection.

4.2 Integration with Cognitive Skill Acquisition Frameworks

Our empirical findings provide computational evidence supporting the phase transitions in skill acquisition proposed by Tenison et al. (2016). The distinct correlation patterns between certain skills and reaction times across learning encounters suggest qualitative shifts in cognitive processing rather than continuous improvements within a single processing approach. Specifically, the stable correlation between Skill 3 and response times across all encounters ($r = 0.56$ to 0.84) parallels the computation-dominant first phase identified through fMRI analysis by Tenison et al. (2016), where problem-solving involves a sequence of arithmetic computations. The decreasing correlations observed for Skills 1 and 2 in later encounters may reflect transitions to the retrieval-based second phase, where answers are retrieved directly without calculation, and potentially to the automatic response third phase, where problems are recognised as unified patterns.

This interpretation gains further support from the memory decay rate analysis. The fact that Skill 3 maintains strong positive correlations with memory decay rates across all Level 3 encounters suggests it captures elements of computational processing that remain cognitively demanding despite practice. In contrast, the weaker and more variable correlations for Skills 1 and 2 may indicate these skills increasingly reflect retrieval-based or automatic processing that becomes less susceptible to forgetting with continued practice, consistent with Anderson (1982) progression from declarative to procedural knowledge.

From the perspective of Taatgen and Blankestijn (2024) hierarchical skill framework, our three-skill configuration may represent distinct levels in the combinatorial structure of multiplication knowledge. Skill 3, with its consistent association with longer response times, potentially represents lower-level computational components such as repeated addition or counting. Skills 1 and 2, with their more variable and generally decreasing correlations with response times, might capture higher-level combinatorial skills such as derived-fact strategies or direct retrieval. The differential impact of practice on these skills aligns with the Taatgen and Blankestijn (2024) proposition that mathematical competencies manifest as transferable skills that can be hierarchically organised and differentially affected by learning experiences.

While our current data and analysis methods cannot definitively establish the neurological or cognitive mechanisms underlying these skill patterns, the alignment with established theoretical frameworks strengthens the interpretation that GraafTel is capturing meaningful cognitive transitions rather than arbitrary statistical patterns. Future research incorporating neuroimaging techniques similar to those employed by Tenison et al. (2016) or more direct strategy assessments could further validate these theoretical connections.

4.3 Strategy Application in Digital Learning Environments

An important methodological consideration emerges when comparing our findings to the strategy framework proposed by Siegler (1988). Siegler's original research examined strategy use in contexts where students typically had access to pen and paper, providing external memory support and calculation aids. In contrast, the TafelTrainer digital environment presents multiplication problems without these physical tools, potentially altering the cognitive demands and the specific implementations of strategies.

This environmental difference may have significant implications for how we interpret the observed skill patterns. In a digital-only environment, counting-based strategies such as repeated addition (e.g., solving 3×4 as $4 + 4 + 4$) might impose greater working memory demands since students cannot offload intermediate calculations to paper. This increased cognitive load could manifest in longer response times and potentially higher error rates for problems that would typically be solved through calculation rather than retrieval. The consistently strong correlation between Skill 3 and response times may reflect this heightened cognitive demand associated with mental calculation without external support.

Furthermore, the absence of physical calculation aids might accelerate the transition from calculation-based to retrieval-based strategies. When calculation is more cognitively demanding due to the lack of external support, the relative advantage of direct retrieval increases, potentially incentivising students to develop stronger associative connections between problems and answers. This environmental pressure could explain the patterns observed in Skills 1 and 2, where correlations with response times decreased in later encounters, suggesting a shift toward more efficient retrieval strategies.

From a methodological perspective, these considerations highlight the importance of contextualising strategy use within specific learning environments. The strategies identified by Siegler (1988) remain conceptually relevant, but their specific implementations and cognitive demands may differ substantially across physical and digital contexts. Future research examining strategy use in digital learning environments should explicitly account for these differences, perhaps by comparing strategy deployment across different presentation modalities or by incorporating think-aloud protocols that can reveal how students adapt strategies to environment-specific constraints.

4.4 Cognitive Mechanisms and Response Patterns

The correlation patterns between GraafTel skill probabilities and reaction times provide insights into potential cognitive mechanisms underlying multiplication fact learning. The consistent correlation between Skill 3 and response times ($r = 0.56$ to 0.84) across learning encounters suggests this skill may correspond to calculation-based processes that remain time-consuming regardless of practice. According to Siegler (1988), calculation-based strategies such as repeated addition or counting impose greater cognitive demands and consistently require more processing time than direct retrieval. The stability of Skill 3's correlation with response times aligns with this theoretical prediction, suggesting this skill captures cognitive processes that remain fundamentally time-dependent even with practice.

In contrast, Skills 1 and 2 showed more variable and generally decreasing correlations with response times across learning encounters, particularly in Level 2. This pattern is consistent with the development of retrieval-based strategies, which become faster and more efficient with practice as associative connections strengthen. The differential patterns observed between skills provide preliminary evidence for the computational differentiation between calculation-based and retrieval-based processes, though our current analysis cannot definitively establish the specific cognitive mechanisms represented by each skill.

The negligible correlations observed at the student level for memory decay rates might indicate that strategy selection operates primarily at the problem level rather than being determined by individual student characteristics. This aligns with Siegler (1988)'s finding that strategy choice is highly adaptive and problem-specific, with students employing different strategies for different problems based on factors such as problem difficulty and prior success. This pattern suggests that cognitive mechanisms underlying multiplication fact learning may be more significantly influenced by item-specific characteristics than by general learner traits, though this interpretation requires further validation through research that explicitly measures strategy use alongside computational modelling.

4.5 Educational Implications

Our findings suggest several concrete implications for educational practice, though these should be considered preliminary given the methodological limitations discussed later. First, the computational differentiation between calculation-based and retrieval-based strategies supports instructional approaches that deliberately cultivate both procedural fluency and fact retrieval, rather than exclusively emphasising memorisation. Educational interventions might benefit from explicitly teaching students when to apply different strategies based on problem characteristics and learning stage.

Second, the distinct correlation patterns between specific multiplication facts suggest potential improvements in how multiplication facts are sequenced and taught. Traditional instructional sequences often present multiplication tables in numerical order, but our findings indicate that cognitive learning pathways might follow different organisational principles. For example, the strong correlations observed between certain skill probabilities and response times might inform more effective grouping for instruction, such as teaching facts that share similar strategic approaches together rather than adhering strictly to numerical sequence.

Third, our findings highlight the importance of adapting instructional approaches to specific learning environments. The differences between digital and traditional learning contexts, particularly regarding the availability of external memory aids, suggest that digital multiplication practice might benefit from several adaptations. Initial scaffolding could gradually fade as students develop greater working memory capacity and retrieval proficiency. Adaptive difficulty progression might account for the potentially increased cognitive load of mental calculation without external support. Explicit strategy instruction could be tailored to the constraints of digital environments, while customised practice schedules might prioritise facts based on empirically derived skill hierarchies rather than conventional sequencing.

Educational technology developers might implement these insights by incorporating more sophisticated adaptive algorithms that model individual strategy development trajectories and adjust practice accordingly. For instance, TafelTrainer's existing adaptive scheduling could be enhanced to differentiate between calculation-dominant and retrieval-dominant phases, providing appropriately tailored practice experiences as students progress through these stages.

These insights also have implications for the design of adaptive learning systems like TafelTrainer. The three-level structure implemented in Tafel-

Trainer aligns conceptually with the three learning phases identified by Tenison et al. (2016), suggesting an effective approach for facilitating the transition from computation to retrieval to automaticity. However, our findings indicate that adaptation parameters could be further refined to account for skill-specific learning trajectories. For example, facts that heavily rely on Skill 3 (potentially computation-based processing) might benefit from different practice schedules than those primarily involving Skills 1 and 2 (potentially retrieval-based processing). Additionally, the GraafTel approach to identifying skill hierarchies described by Taatgen and Blankestijn (2024) could complement the memory model currently used in TafelTrainer (Iancu et al., 2024), potentially leading to more sophisticated adaptation algorithms that account for both forgetting curves (van Rijn et al., 2009) and skill dependencies in scheduling practice opportunities.

4.6 Limitations

Several important limitations of the current study constrain the interpretation of our findings and must be acknowledged when considering their implications.

First, our computational approach provides indirect evidence of strategy use but cannot directly observe students' thought processes or conscious strategy choices. The inference of strategy types from performance patterns, while theoretically grounded, remains speculative without direct verification. The GraafTel modelling approach identifies distinct skill patterns that correlate meaningfully with performance metrics, but the mapping between these computational constructs and specific cognitive strategies relies on theoretical interpretation rather than direct measurement. This limitation is particularly relevant when interpreting the specific cognitive processes underlying the three identified skills, as these interpretations remain provisional without converging evidence from complementary methodologies.

Second, the measurement of response times in digital environments introduces potential confounds that are difficult to disentangle from strategy-related variations. Typing proficiency, interface familiarity, and technological distractions might all influence response latencies independently of mathematical strategy selection. The TafelTrainer platform, while providing ecological validity through its implementation in authentic educational settings, lacks the experimental control necessary to isolate strategy-specific effects from these potential confounding factors. This limitation affects the interpretation of response time patterns, particularly when making fine-grained distinctions between different strategy types based on temporal metrics.

Third, the student-level analysis of memory decay rates showed negligible correlations across skills, which might indicate limitations in our measurement approach or suggest more complex interactions than our model captured. The adaptive nature of the learning system, while educationally valuable, presents analytical challenges for isolating stable individual differences. The continuous adjustment of practice scheduling based on performance potentially masks individual variation in learning trajectories, making it difficult to distinguish between strategy-related and method-related effects. This limitation affects the generalisability of our findings regarding individual differences in strategy development patterns.

Fourth, the ecological validity gained through our naturalistic study design introduces variability in implementation conditions across participating schools and classrooms. Teachers retained autonomy over scheduling and integration of the TafelTrainer application, resulting in considerable variation in usage patterns across participants. This implementation variability, while representative of real-world educational technology deployment, creates analytical challenges when attempting to isolate specific learning effects from contextual factors. The varying exposure levels across participants potentially influences the observed skill patterns, limiting the precision with which developmental trajectories can be mapped across the study population.

4.7 Future Research Directions

Building upon the findings and limitations of the current study, several promising research directions emerge that could further elucidate the cognitive mechanisms underlying multiplication fact acquisition and inform more effective educational interventions.

Strategy verification studies could combine computational modelling with direct strategy assessment methods, such as verbal protocols or strategy choice questionnaires. By explicitly asking students to articulate their solution processes during or immediately after problem-solving, researchers could establish more direct links between the computational skill patterns identified through Graaf-Tel and specific cognitive strategies. Such multi-method approaches would strengthen the theoretical interpretation of skill patterns while potentially revealing additional strategies or hybrid approaches not captured by performance metrics alone. Particularly valuable would be studies examining how strategy verbalisation corresponds to skill probabilities across different learning stages, providing empirical validation for the hypothesised mapping between computational skills and cognitive processes.

Longitudinal investigations examining how strategy patterns evolve over extended time periods and persist after instruction concludes would address the temporal limitations of the current study. Following students over multiple academic terms while periodically reassessing both performance metrics and explicit strategy use would reveal whether the transitions we observed between calculation-based and retrieval-based approaches represent stable changes in cognitive processing or temporary adaptations to specific learning contexts. Such studies could also examine how strategy development trajectories relate to broader mathematical competencies, potentially establishing predictive relationships between early strategy flexibility and later mathematical achievement.

Comparative analyses across learning environments could directly contrast strategy development in digital versus traditional learning contexts. Systematically varying the availability of external memory aids (e.g., paper and pencil) within a controlled experimental design would help isolate the specific effects of environmental constraints on strategy selection and development. This research could identify environment-specific affordances that influence cognitive processing, informing more targeted instructional approaches that account for the particular cognitive demands of digital learning environments. Such studies would be particularly valuable for understanding how the findings from traditional strategy research (e.g., Siegler (1988)) translate to increasingly prevalent digital learning contexts.

Individual difference analyses could examine how learner characteristics such as working memory capacity, processing speed, or prior mathematical achievement influence strategy development trajectories. By combining psychometric assessment of these cognitive factors with longitudinal tracking of strategy development, researchers could identify whether certain cognitive profiles predict distinct patterns of strategy acquisition or particularly benefit from specific instructional approaches. This research would contribute to more personalised educational interventions that account for individual cognitive constraints and affordances in the learning process.

Educational intervention studies could test instructional approaches informed by our findings, such as sequencing multiplication facts based on empirically derived skill hierarchies rather than conventional ordering. Experimental comparisons between traditional instructional sequences and cognitively informed alternatives would directly assess whether instruction aligned with computational skill structures yields superior learning outcomes. Such studies could also examine whether different instructional sequences are differentially effective for learners with varying cognitive profiles, potentially leading to adaptive educational approaches that optimise learning pathways based on individual characteristics and developmental trajectories.

References

Abdelrahman, G., Wang, Q., & Nunes, B. (2023). Knowledge tracing: A survey. *ACM Computing Surveys*, *55*. https://doi.org/10.1145/3569576

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological review*, *89*(4), 369.

Campbell, J. I., & Thompson, V. A. (2012). Retrieval-induced forgetting of arithmetic facts. *Journal of Experimental Psychology: Learning Memory and Cognition*, *38*. https://doi.org/10.1037/a0025056

Das, J. P., & Janzen, C. (2004). Learning math: Basic concepts, math difficulties, and suggestions for intervention. *Developmental Disabilities Bulletin*, *32*.

Dehaene, S. (2011). *The number sense: How the mind creates mathematics*. OUP USA.

del Carmen Chamorro, M. (2021). Can we explain students' failure in learning multiplication? In M. Isoda & R. Olfos (Eds.), *Teaching multiplication with lesson study: Japanese and ibero-american theories for international mathematics educationforeword by abraham arcavi* (pp. 265–289). Springer International Publishing. https://doi.org/10.1007/978-3-030-28561-6_11

Elo, A. E. (1978). *The rating of chess players, past present*.

Gilmore, C. (2023). Understanding the complexities of mathematical cognition: A multilevel framework. *Quarterly Journal of Experimental Psychology*, *76*. https://doi.org/10.1177/17470218231175325

Iancu, S. D., Braam, M., Mccabe, N., Wilschut, T., Rijn, H. V., & van der Velde, M. (2024). An adaptive learning system for stepwise automatisation of multiplication facts in primary education.

Kaskens, J., Goei, S. L., Luit, J. E. V., Verhoeven, L., & Segers, E. (2022). The roles of arithmetic fluency and executive functioning in mathematical problem-solving. *Elementary School Journal*, *123*. https://doi.org/10.1086/721771

Kingma, D. P., & Ba, J. (2017). Adam: A method for stochastic optimization. https://arxiv.org/abs/1412.6980

Klinkenberg, S., Straatemeier, M., & Maas, H. L. V. D. (2011). Computer adaptive practice of maths ability using a new item response model for on the fly ability and difficulty estimation. *Computers and Education*, *57*. https://doi.org/10.1016/j.compedu.2011.02.003

Lemaire, P., & Siegler, R. S. (1995). Four aspects of strategic change: Contributions to children's learning of multiplication. *Journal of experimental psychology: General*, *124*(1), 83.

Sense, F., Behrens, F., Meijer, R. R., & van Rijn, H. (2016). An individual's rate of forgetting is stable over time but differs across materials. *Topics in Cognitive Science*, *8*. https://doi.org/10.1111/tops.12183

Siegler, R. S. (1988). Strategy choice procedures and the development of multiplication skill. *Journal of Experimental Psychology: General*, *117*(3), 258–275. https://doi.org/10.1037/0096-3445.117.3.258

Siegler, R. S. (2009). Improving the numerical understanding of children from low-income families. *Child Development Perspectives*, *3*(2), 118–124.

Taatgen, N. A. (2013). The nature and transfer of cognitive skills. *Psychological review*, *120*. https://doi.org/10.1037/a0033138

Taatgen, N. A., & Blankestijn, J. (2024). Data-driven cognitive skills with an application in personalized education.

Tenison, C., Fincham, J. M., & Anderson, J. R. (2016). Phases of learning: How skill acquisition impacts cognitive processing. *Cognitive psychology*, *87*, 1–28.

van Rijn, H., van Maanen, L., & van Woudenberg, M. (2009). Passing the test: Improving learning gains by balancing spacing and testing effects. *Proceedings of ICCM 2009 - 9th International Conference on Cognitive Modeling*.

van der Velde, M. (2024, February). Building tafeltrainer #3: How we designed an adaptive algorithm for learning multiplication facts.

A Appendix

The dataset includes 359,950 rows, capturing detailed records of individual practice sessions and trials and comprising several key columns, each contributing information helpful for analysing the learning process:

- **user_id:** An anonymised identifier for each student, ranging from 17 to 832.
- **session_id:** A unique identifier for each practice session, recorded as a UUID string (e.g., `"407f83a8-fde2-4a44-b7be-32b16cd38ed3"`), which allows linking of multiple trials within the same session.
- **level:** The adaptive level of the session, with values of 1, 2, or 3. Each level corresponds to a specific type of learning approach:
- **algorithm:** The practice algorithm used in the session, corresponding to the session level:
  - **none:** used in Level 1.
  - **accuracy only:** used in Level 2.
  - **accuracy + rt:** used in Level 3.
- **fact_id:** An integer identifying each multiplication fact, where `1` corresponds to `"1 x 1"`, `2` corresponds to `"1 x 2"`, and so on.
- **cue_text:** A human-readable representation of the multiplication fact (e.g., `"1 x 1"`).
- **answer:** The correct answer to the multiplication fact (e.g., `1` for `"1 x 1"`).
- **given_response:** The response provided by the student during the trial (e.g., `2`).
- **correct:** A Boolean value (`TRUE` or `FALSE`) indicating whether the student's response matched the correct answer.
- **reaction_time:** The time elapsed in milliseconds from the start of the trial to the student's first keypress, available for Levels 2 and 3.
- **alpha:** A "speed of forgetting" parameter, representing the rate at which a student is likely to forget specific facts, available for Levels 2 and 3 trials.
- **presentation_start_time:** The Unix timestamp indicates the onset of each trial, recorded in milliseconds.
- **session_time:** The time elapsed since the start of the session, recorded in milliseconds.

We preprocessed the dataset to standardise and clean the data for analysis. The `correct` column, which was a Boolean value indicating whether a response was correct (`TRUE/FALSE`), was converted to numeric values (`1` for correct and `0` for incorrect) to facilitate statistical computations and compatibility with the GraafTel system. Additionally, the `cue_text` column required cleaning to ensure consistency in its format. In some instances, strings of multiplication problems such as `6 x 3` included additional characters, for example, `"6+x+3"`. To standardise the column, these plus signs (`"+"`) were replaced with spaces, ensuring a uniform format across all entries.

Two new columns, `multiplier` and `multiplicand`, were derived from the cleaned `cue_text` column. These variables extracted the operands of the multiplication (e.g., `"6 x 1"` produced `multiplier = 6` and `multiplicand = 1`) and enabled systematic grouping and analysis of performance by specific multiplication components.

Each trial was assigned an encounter number (`encounter_num`) based on its sequential order within a practice session. This procedure made it easier to divide the dataset into specific encounter types.

For GraafTel compatibility, the dataset was refined to include only the required columns: `Student ID`, `Question ID`, and `Accuracy`. Accuracy values were ensured to remain between 0 and 1, adhering to GraafTel's specifications. The `cue_text` column, which provides a short, descriptive label of each multiplication fact (e.g., `6 x 1`), was used as the `Question ID`, making the data interpretable during GraafTel analysis. Column headers were removed from the exported files, and separate CSV files were created for each adaptive level (1, 2, and 3) and encounter type (first, middle, and last).

B Appendix

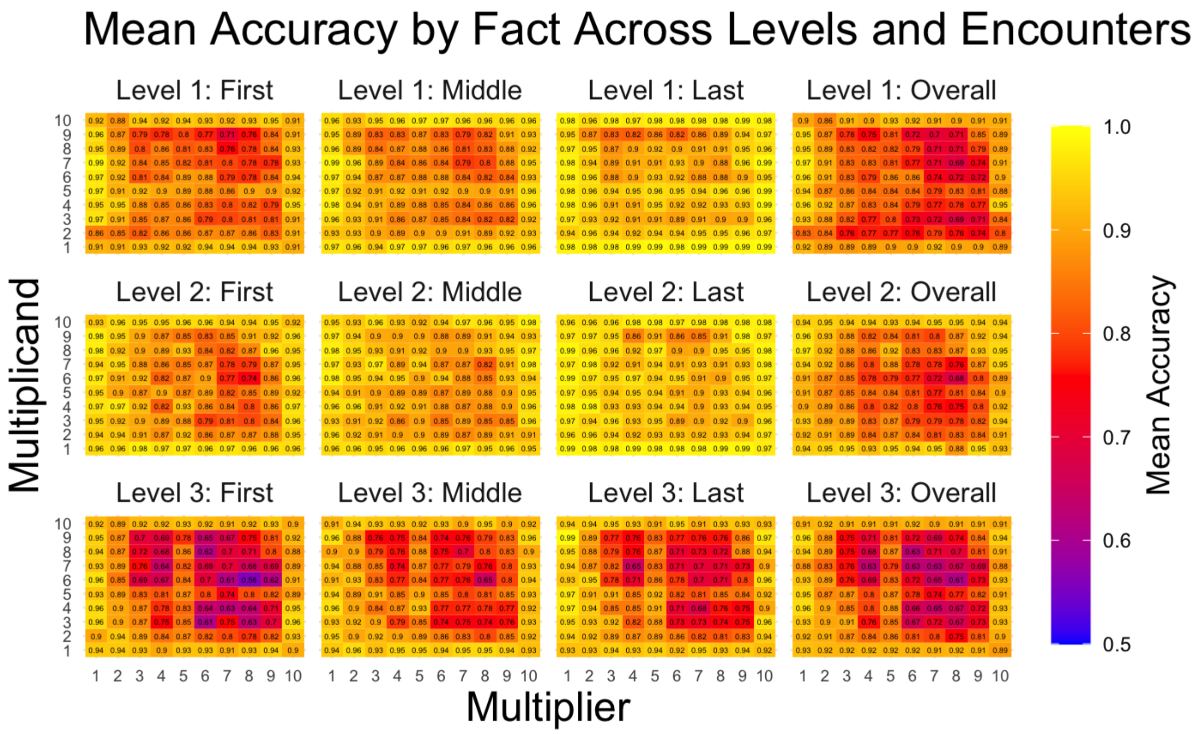

Figure B.1: Mean accuracy across levels and encounters for each multiplication fact. The heatmap matrix displays mean accuracy for multiplication facts, grouped by multiplicand (y-axis) and multiplier (x-axis), across three levels of learning (L1, L2, and L3) and three encounter positions (First, Middle, Last), as well as an overall summary for each level. Warmer colors (yellow) indicate higher accuracy, while cooler colors (red to blue) indicate lower accuracy. Accuracy is generally highest for facts involving 1s, 2s, 5s, and 10s, while mid-range facts show more variation, particularly in Level 3, where some accuracy rates drop below 0.6.