Automatic Modeling and Analysis of Students' Problem-Solving Handwriting Trajectories

Zhonghua Sheng[1*], Shuyu Shen[2*], Leixian Shen[1(✉)], Qiqi Duan[2], Nan Tang[1,2], Pan Hui[1,2], Huamin Qu[1], and Yuyu Luo[1,2]

The Hong Kong University of Science and Technology, China
The Hong Kong University of Science and Technology (Guangzhou), China
lshenaj@connect.ust.hk

Abstract. Understanding students' cognitive patterns in problem-solving is crucial for personalized education, yet traditional methods struggle to effectively capture and analyze these patterns. This paper presents CogChain, a novel method that synergistically combines digital pen technology with Multi-modal Large Language Models (MLLMs) to automatically construct students' logic chains during examinations. We collected a comprehensive dataset of 87,679 handwriting trajectories in mathematics, physics, and chemistry from 25 real-world high school students. Based on the constructed logic chains of students, we conduct an in-depth analysis across three dimensions: solution, time, and course, revealing a set of findings about their problem-solving behaviors. (1) Solution Dimension: We identify four distinct solution trajectory patterns, showing that moderate-complexity approaches achieve the highest accuracy. (2) Time Dimension: We uncover three types of time-allocation patterns, showing that students who allocate more time to structured reasoning achieve higher accuracy, whereas those who prioritize writing speed tend to perform worse. (3) Course Dimension: Different subjects require distinct problem-solving and time management strategies, with mathematics benefiting from step-by-step derivation, physics from visual reasoning, and chemistry from quick solutions. These insights provide valuable guidance for developing personalized teaching strategies.

Keywords: Handwriting Trajectory · Multi-modal Large Language Model · Logic Chains · Cognitive Patterns.

1 Introduction

Examinations are essential for assessing student learning, with students documenting their problem-solving processes and answers on paper, which educators evaluate for accuracy and completeness [2, 24]. These written records contain rich information about students' problem-solving trajectories, offering insights

Z. Sheng and S. Shen contributed equally to this work.

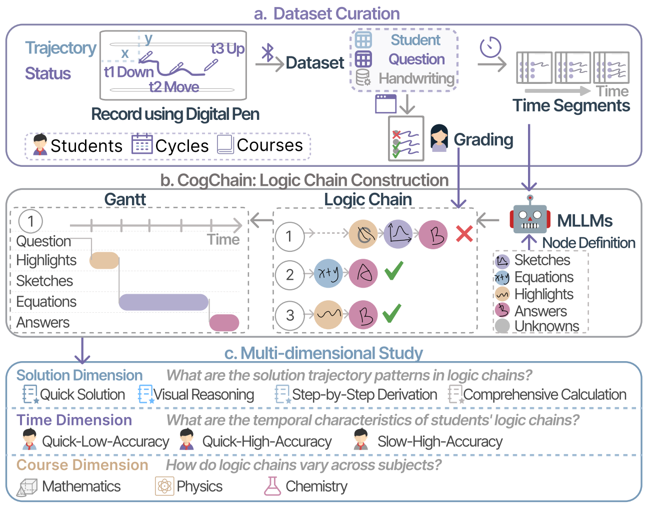

Fig. 1. Overview of three key components: (a) Data collection via digital pens. (b) CogChain: MLLM-based logic chain construction from handwriting trajectories. (c) Multi-dimensional analysis of logic chains across solution, time, and course dimensions.

into students' thinking patterns and cognitive modes [18, 19]. For example, some students tend to use sketches to draw auxiliary diagrams, visually constructing their solution approaches, while others directly enter the equation phase, focusing on formula derivation and calculations.

However, faced with large numbers of students and complex test questions, it is challenging to analyze each student's entire solution process individually, and students often struggle to accurately recall their thought processes during problem solving [1, 27, 5]. To tackle this challenge, analyzing handwriting trajectories can reconstruct the sequential steps linking initial understanding to final solutions, forming what we call ***logic chains*** [8, 15, 7, 4]. These chains represent the structured progression of reasoning and actions taken to solve a problem, revealing two key aspects of students' cognitive patterns: solution trajectory patterns that show how students construct their solutions (e.g., through visual aids or direct equation derivation), and time allocation patterns that reflect how students manage their problem-solving process. Understanding these cognitive patterns provides opportunities for more targeted personalized guidance and helps educators optimize teaching strategies [14, 3].

Capturing Handwriting Trajectories with Digital Pens. Digital pens, as innovative data collection tools, can record real-time trajectory coordinates (x, y) and temporal information (time, status) of pen movements on paper during students' problem-solving processes, as shown in Figure 1(a) and Figure 2. These raw data consist merely of spatiotemporal coordinate points that require alignment with answer sheets for meaningful analysis. However, without proper processing, it remains challenging for educators to extract meaningful insights from this rich dataset [28, 32].

Logic Chain Modeling. To address the challenge of extracting insights from raw trajectory data, we leverage the recent advances in Multimodal Large Language Models (MLLMs) [12, 23]. Their ability to process multimodal inputs, including handwritten notes, makes them a promising tool for analyzing students' problem-solving handwriting trajectories. Building on this, we propose CogChain, an MLLM-based method to automatically analyze students' time-sequential handwritten trajectories with digital pens (see Figure 1(b)). CogChain extracts spatiotemporal features, visualizing them as logic chains that can be transformed into Gantt charts [9], illustrating different cognitive phases (e.g., sketch drawing and thinking processes).

Based on the extracted logic chains, we further analyze students' cognitive patterns in three dimensions: solution, time, and course (Figure 1(c)).

- ***Q1:*** *What are the solution trajectory patterns in logic chains (Solution Dimension)?* We aim to explore the relationships between students' solution trajectory patterns and knowledge topics to reveal trajectory characteristics.
- ***Q2:*** *What are the temporal characteristics of students' logic chains (Time Dimension)?* We aim to explore the relationships between students' time allocation and solution trajectory patterns to reveal temporal characteristics.
- ***Q3:*** *How do logic chains vary across subjects (Course Dimension)?* We aim to compare solution trajectory patterns across different courses to identify cognitive differences.

Our contributions are summarized as follows:

- **Dataset Curation.** We developed a handwriting dataset with digital pens, containing two months of records from 25 high school students (see Figure 1(a)). The dataset comprises test papers in mathematics, physics, and chemistry, with a total of 87,679 handwriting trajectories.
- **CogChain: Automatic Logic Chain Construction.** We propose CogChain, an MLLM-based method to automatically generate logic chains from handwriting trajectories (see Figure 1(b)). Logic chains display temporal distribution and relationships between trajectories, providing a new method for understanding students' problem-solving behaviors.
- **Multi-Dimensional Study of Logic Chains.** We conducted an in-depth analysis of logic chains in three dimensions: solution, time, and course (see Figure 1(c)). By analyzing logic chains, we explore the students' solution trajectory patterns and time management patterns across different subjects, providing valuable insights into students' cognitive patterns.

2 Related Work

2.1 Digital Pen for Education

Digitalizing student handwriting can significantly enhance grading efficiency and has the potential to enable intelligent, personalized education [11]. Compared to handwritten word recognition systems [13], digital pens offer several advantages: they capture temporal information [29], facilitate intuitive interaction [28], and allow for fine-grained analysis of raw data [10, 32, 25, 17, 31, 6]. For instance, Yan [31] demonstrated that smart paper-pen technology improves classroom interactions by making them more efficient. Guo et al. [10] developed a learning platform utilizing digital pens, incorporating an analytical model to assess learners' engagement with assignments based on validated indicators. Similarly, Siddiqui and Muntjir [25] introduced a smart study platform integrating digital pens and paper to support students in reading, writing, and solving classroom exercises. Despite these advantages, there remains a lack of digital pen handwriting trajectory datasets, and a shortage of fine-grained analyses of these trajectories. This paper addresses this by constructing a real-world dataset and leveraging MLLMs to analyze cognitive patterns from handwriting trajectories.

2.2 Student Logic Chain Analysis

Understanding and modeling students' logic chains during problem-solving has attracted significant attention. Wang and Chiew [30] highlighted problem-solving as a core cognitive function intertwined with abstraction, learning, decision-making, inference, and synthesis. They further captured these interactions to study the complexities of student cognition. Recent AI advancements have introduced techniques like chain-of-thought prompting [26], which train models to use step-by-step reasoning to handle complex tasks with greater accuracy. This aligns with the goal of modeling student logic chains, as it emphasizes the decomposition of complex problems into manageable steps. Furthermore, the use of digital tools in education has opened new avenues for capturing and analyzing student problem-solving processes. Digital pens, for instance, can record real-time trajectory coordinates and temporal information during students' problem-solving activities. This data can be analyzed to reconstruct the sequence of cognitive steps taken by students, providing insights into their logic chains [10, 25]. Despite these advances, current methods face key limitations of insufficient granularity in capturing transitions between cognitive phases [20, 16]. To address these gaps, we propose leveraging MLLMs to reconstruct fine-grained, time-sequential logic chains from digital pen handwriting trajectories.

3 Dataset Curation

We collected data from a secondary school class with 25 students (10 male and 15 female, aged 17-19). The examinations were conducted under standard conditions, following students' training on digital pen usage. Students then completed timed exams using specially designed answer sheets embedded with coordinates, enabling precise tracking of pen movements (see Figure 2(a)). Their problem-solving experience remained identical to traditional paper-based exams, while we collected handwriting trajectory data via the pen's Bluetooth functionality.

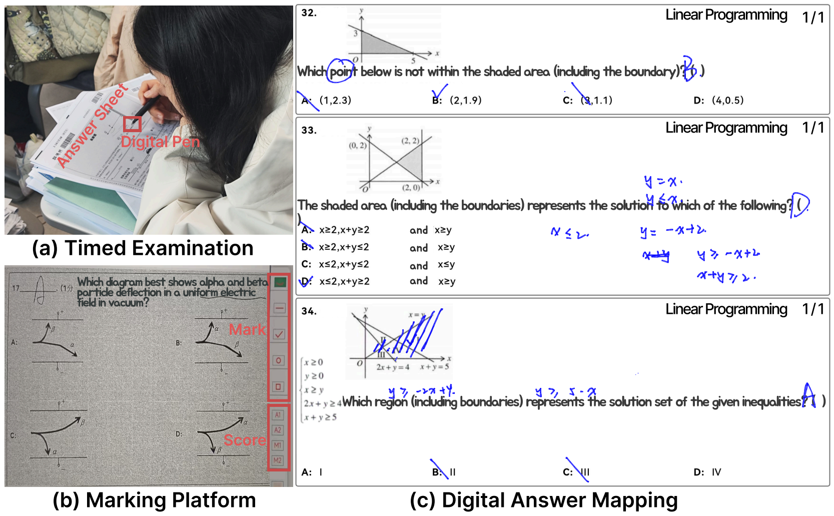

Fig. 2. Example of handwriting samples in the dataset

Over a two-month period, we collected 37 examination papers across mathematics, physics, and chemistry. Each pen stroke was recorded with state information (pen down, movement, and pen up), spatial features (x- and y-coordinates), and timestamps. We define a trajectory as the complete sequence of pen movements from pen-down to pen-up, including all intermediate positions and their spatial coordinates. This process resulted in 87,679 trajectory segments. We further preprocessed these trajectories by mapping the coordinates to standard answer sheets. Figure 2(c) illustrates how a student solved three linear programming problems. For Q32, the student circled the question stem, eliminated incorrect options, and wrote 'B'. For Q33, the student wrote equations, crossed out wrong choices, and selected 'D'. For Q34, the student wrote equations, created graphs, and chose 'A' after eliminating incorrect options. Educators can grade and recorde results through a marking platform (see Figure 2(b)).

The dataset comprises three main components: student information with unique identifiers, question information including correct answers and knowledge topics, and handwriting data containing these detailed pen stroke records (see Figure 1(a)). This dataset enables in-depth analysis of problem-solving patterns, cognitive strategies, and personalized learning insights.

4 CogChain: Automatic Logic Chain Construction

A logic chain represents a student's structured problem-solving process, comprising writing phases (e.g., equations, sketches), thinking gaps (pauses indicating reasoning or analysis), and temporal organization (the sequential flow reflecting logic and time allocation). To gain insights into students' cognitive patterns, we propose CogChain, a method that leverages MLLMs to automatically construct and visualize logic chains from students' raw handwritten trajectory data. As

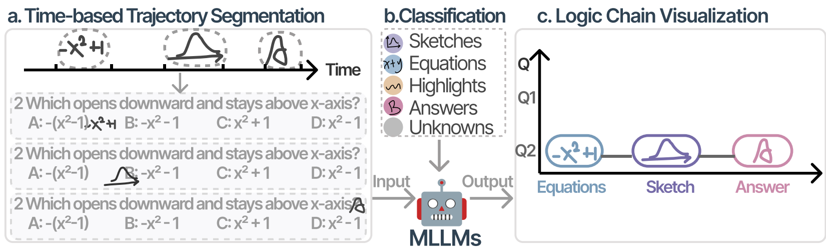

Fig. 3. CogChain: time-sequential handwriting-based logic chain construction.

shown in Figure 3, CogChain comprises three modules: time-based trajectory segmentation, trajectory classification, and logic chain visualization.

4.1 Time-based Trajectory Segmentation

To generate students' logic chains more accurately, we first adopt a time-based trajectory segmentation method to process large amounts of the handwritten trajectories (see Figure 3(a)). Considering that inputting all trajectories at once would lose temporal sequence information and excessive trajectory might result in a loss of accuracy, we employ a heuristic grouping method by setting a time threshold (2s) to segment students' writing behavior. When the time interval between adjacent trajectories is less than the 2s threshold, we consider it a continuous writing phase. For example, when a student continuously writes a sequence of a 'x' letter within the 2-second threshold, these strokes are grouped as a single writing behavior, preserving the natural flow of writing. During this phase, we group these continuous pen trajectories together and record their writing duration. When the time interval exceeds the 2s threshold, we identify it as a student's thinking pause phase and record this period as a thinking gap.

As shown in Figure 3(a), this method segments the student's trajectory into three distinct writing groups, each associated with a writing duration, with two thinking gaps in between. To facilitate logic chain construction, we map these groups onto answer sheets in chronological order. Figure 3(a) illustrates how these three writing groups are mapped onto three separate answer sheets.

4.2 Handwritten Trajectory Classification

Based on a comprehensive analysis of our dataset, we categorized handwritten trajectory groups into five distinct types that constitute the writing phases of students' logic chains (see Figure 3(b)): *Sketches*: Mathematical diagrams and graphical elements (e.g., coordinate axes, numerical lines); *Equations*: Mathematical expressions serving as auxiliary calculations or final answers; *Highlights*: Marking symbols including strikethroughs and underlines; *Multiple-choice answer*: Letter-based responses (e.g., A, B, C, D) enclosed in brackets; and *Unknowns*: Unidentifiable marks or irrelevant content (e.g., doodles, stray marks).

The classification process leverages MLLMs to analyze segmented handwriting. As students often switch between questions, handwriting may inadvertently overlap with neighboring questions. To address this, our MLLMs perform dual classification, identifying both handwriting type and corresponding question. By analyzing spatial distribution and content features of trajectory images mapped to answer sheets, MLLMs ensure accurate categorization.

4.3 Logic Chain Visualization

We represent the logic chain of students' problem-solving processes using Gantt charts (see Figure 3(c)). These charts integrate three key data components: writing durations and thinking gaps from trajectory segmentation, and handwriting classification results from MLLMs. The Gantt chart's horizontal axis represents the temporal progression, while the vertical axis denotes different questions. Writing phases are color-coded based on their types (e.g., equation, sketch, answer), and thinking gaps appear as grey regions. As an example of how our visualization method captures problem-solving processes, consider a student's solution to the question *"Which opens downward and stays above the x-axis?"* shown in Figure 3. The Gantt chart reveals a structured sequence: the student began by noting relevant formulas, paused for strategic thinking, constructed function coordinates, and concluded by marking 'A' as the final answer. This visualization approach effectively captures both the sequential nature of students' solution trajectory patterns and their time allocation patterns.

5 Multi-Dimensional Study of Logic Chains

Through the automatic logic chain construction method described in Section 4, we analyze 87,679 digital pen trajectory segments collected from 37 examination papers, thus identifying 8,304 trajectory groups utilizing CogChain. By associating these trajectory groups with their corresponding questions, we obtain 1,765 complete problem-solving processes, where each process might contain multiple trajectory groups representing different phases of solving a question. We further conduct an analysis of students' cognitive patterns based on these logic chains across three dimensions: solution, time, and course.

5.1 Solution Dimension

Experiment Design. To analyze students' problem-solving trajectory characteristics, we first define solution trajectory patterns as clusters of similar writing phase sequences within logic chains for individual questions. To identify these patterns, we perform cluster analysis on 1,765 problem-solving processes by extracting two features from each logic chain: the total number of trajectory groups and the time-based distribution of different trajectory types defined in Section 4.

Using these features, we employ K-means clustering with the elbow method to determine the optimal number of clusters. Through this analysis, we identified four typical solution trajectory patterns:

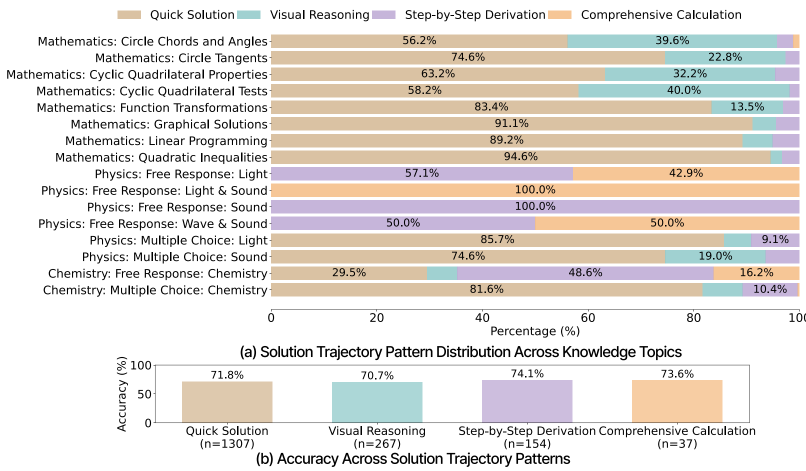

Fig. 4. Solution trajectory patterns in logic chains.

- *Quick Solution*: A concise problem-solving approach with few trajectory groups (avg. 2.8) where students predominantly use highlighting (49.1%) for multiple-choice questions. The key feature is minimal calculation (avg. 0.6 equations) and highlighting use (avg. 0.8), achieving 71.8% accuracy.
- *Visual Reasoning*: A solution method with moderate trajectory groups (avg. 6.7) dominated by sketches (54.4%). The distinguishing characteristic is the incorporation of graphical elements (avg. 2.8 sketches), such as coordinate systems and function plots, reaching 70.7% accuracy.
- *Step-by-Step Derivation*: A solution process with more trajectory groups (avg. 11.2) combining equations (44.9%) and highlighting (46.9%). It is characterized by equation derivations (avg. 3.8) and highlighting (avg. 3.9), showing systematic problem-solving logic with 74.1% accuracy.
- *Comprehensive Calculation*: An extensive solution approach with numerous trajectory groups (avg. 29.5) characterized by detailed mathematical derivations (51.4% equations). It features equation derivation (avg. 12.3 equations), maintaining 73.6% accuracy.

Findings. To evaluate the effectiveness of different solution trajectory patterns, we analyze: (1) knowledge topics and question types derived from the question information, and (2) solution accuracy based on educators' grading in the marking platform. Through a comprehensive analysis of these metrics across different solutions trajectory patterns (see Figure 4), we identify several key findings:

Trajectory patterns correspond to knowledge topics and question types. Students demonstrate distinct problem-solving preferences across various knowledge topics and question types as shown in Figure 4(a). In geometry-related content like circle chords and angles and cyclic quadrilateral tests, which are visual and spatial, students tend to use visual reasoning in over 20% of solutions. For algebraic topics such as function transformations and graphical solutions, with numerical and symbolic nature, students predominantly prefer quick solution methods, making up over 80% of solutions. Regarding question types, in free-response questions that demand detailed explanations, students are more inclined to use step-by-step derivation and comprehensive calculation, with the combined proportion of these two methods exceeding 60% across knowledge topics. In contrast, quick solution methods dominate in multiple-choice questions, with each method exceeding 60% of responses. These patterns suggest that students adaptively select solution methods based on both the inherent nature of knowledge topics and the specific requirements of question formats.

Solution complexity does not guarantee accuracy. We find that extensive solution patterns do not necessarily lead to higher accuracy (see Figure 4(b)). While step-by-step derivations (10-15 trajectory groups) show the highest accuracy at 74.1%, this moderate-complexity approach was only chosen in 154 problem-solving processes. In comparison, quick solutions, which involve just 2-4 trajectory groups with minimal calculations, appear in 1,307 processes and achieve a comparable 71.8% accuracy. Visual reasoning with 5-9 trajectory groups(70.7%, n=267) and comprehensive calculations involving over 15 trajectory groups(73.6%, n=37) show similar performance levels. This suggests that effective problem-solving depends on finding an appropriate balance between detail and conciseness, rather than simply maximizing or minimizing solution trajectories.

5.2 Time Dimension

Experiment design. To deeply understand students' time management during examinations, we first define time allocation patterns as clusters of similar temporal organizations within logic chains for individual students throughout an entire exam. Then, we conduct a cluster analysis of examination time distribution data from 37 completed test papers, which includes several temporal features: average time and standard deviation of thinking gaps and writing durations, exam time management allocations (average time spent during the first and final thirds of the examination), and answer sequence characteristics (reverse order answering ratio). Following the same clustering approach described in Section 5.1, we categorize students into three typical time allocation patterns:

- *Quick-Low-Accuracy*: Characterized by quick writing speeds (avg. 12.1s) with long gap times (avg. 95.8s) and lowest accuracy (45%). Notable for their reverse order answering ratio of 5.3%.
- *Quick-High-Accuracy*: Displaying quick writing speeds (avg. 12.7s) with shortest gap times (avg. 47.2s) and good accuracy (65.5%). Most time-efficient among all clusters with the lowest total solution time (avg. 59.8s).
- *Slow-High-Accuracy*: Showing longer writing times (avg. 42.8s) with long gap times (avg. 106.6s) but highest accuracy (70.3%). Characterized by significant deceleration, with late-stage average time (avg. 249.4s) substantially higher than early stage (avg. 94.3s).

Findings. To identify personalized cognitive patterns in time management, we compare time allocation patterns with both answer accuracy and solution trajectory patterns (derived following Section 5.1), as illustrated in Figure 5.

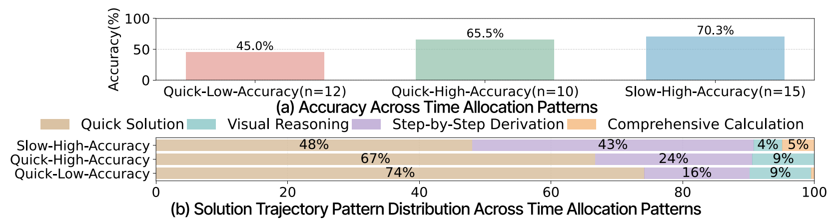

Fig. 5. Time allocation patterns across accuracy and solution trajectory patterns.

Gap time serves different cognitive functions. As shown in Figure 5 (a), distinct gap time patterns across clusters reveal their different roles: the quick-low-accuracy group's long gaps (95.8s) coupled with their low accuracy (45%) may indicate difficulty in problem comprehension despite the extended thinking time, while the slow-high-accuracy group's extended gaps (106.6s) likely reflect deeper analytical processing, contributing to their superior accuracy (70.3%). The quick-high-accuracy group's shorter gaps (47.2s) combined with good accuracy (65.5%) suggest efficient information processing and decision-making capabilities, demonstrating that longer thinking time does not necessarily translate to better performance.

Guiding students toward structured reasoning is essential. As illustrated in Figure 5(b), there are distinct differences in solution approaches across student groups. The quick-low-accuracy group heavily relies on quick solutions (74%), which may contribute to their lower accuracy (45.0%). In contrast, the quick-high-accuracy and slow-high-accuracy groups employ step-by-step derivations more frequently (24% and 43% respectively), achieving notably better results (65.5% and 70.3% accuracy respectively). These findings suggest that encouraging students to spend time identifying key information through highlighting and formulating equations can be beneficial. Particularly, quick-low-accuracy students should be guided to reduce their over-reliance on quick solutions in favor of more systematic approaches, while quick-high-accuracy students could benefit from increased use of step-by-step derivation strategies.

5.3 Course Dimension

Findings. Through a systematic comparison of solution trajectory patterns across different subjects, as shown in Figure 6, we identify differences in answer accuracy, writing duration, and thinking gap (as defined in Section 4.1).

Subject characteristics influence problem-solving strategies and habits. The knowledge structures of different subjects determine their suitable problem-solving strategies (see Figure 6 (a)). Mathematics, with its strict logical nature, is best approached through step-by-step derivation, achieving an average accuracy of 80%. Physics, which involves abstract concepts, benefits from visual reasoning, with this method yielding a high accuracy of 93.8%. Chemistry, characterized by its diverse and complex topics, sees the greatest advantage in quick solution approaches, with an accuracy rate of 80.9%. However, students' actual problem-solving habits do not always align with the optimal strategies. For instance, in

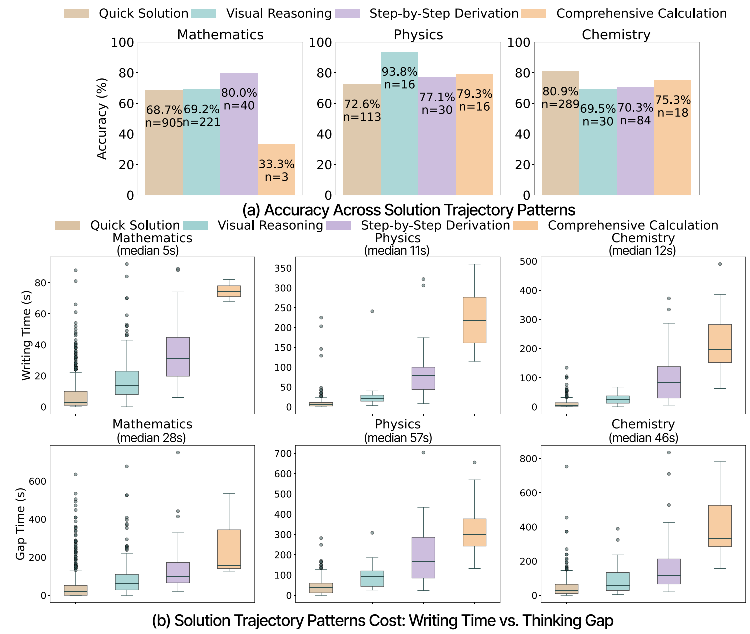

Fig. 6. Solution trajectory patterns across courses: accuracy and time allocation.

mathematics, students tend to favor quick solutions (n=905), but this approach often results in lower accuracy. Conversely, in physics (n=113) and chemistry (n=289), students are more inclined to adopt slower problem-solving strategies. Therefore, students should adjust their problem-solving habits according to the characteristics of each subject to avoid the negative impact of rigid habits on problem-solving effectiveness.

Each subject has its unique time management pattern. As shown in Figure 6(b), mathematics typically has a short writing time (median 5s) and a relatively long thinking time (median 28s), while physics has a writing time of 11s and a thinking time of 57s, and chemistry has a writing time of 12s and a thinking time of 46s. Given that all math questions are multiple-choice, answers are relatively concise, contributing to shorter writing times. Physics and chemistry include both multiple-choice and free-response questions, requiring more detailed solution steps and explanations. Physics shows the longest thinking gap (57s), which may be related to the multi-step problem-solving processes often required in physics problems. Overall, the median of the thinking gap is 4-5 times that of the writing time, indicating a cognitive pattern where analysis and planning dominate the solution process. Notably, quick solutions in all subjects show scattered time distributions, reflecting the inherent instability of rapid problem-solving methods.

6 Discussion

Implication. Our study offers implications across three dimensions:

Cognitive Diagnosis Dimension. Based on our dataset of handwriting trajectories in Section 3, we used CogChain to generate logic chain visualizations (Section 4). Through this analysis, we identified students' cognitive patterns, including problem-solving strategies and time allocation patterns (Section 5). These findings help educators analyze students' problem-solving processes. The cognitive pattern features provide data-driven support for intelligent question bank systems to better assess and adapt to students' learning needs.
Learning Strategy Dimension. Through the analysis of problem-solving data (Section 5), we identified significant individual differences in thinking time allocation, solution selection, and solution accuracy. We constructed comprehensive student cognitive profiles by analyzing trajectory patterns across knowledge points (Section 5.1), examination time management (Section 5.2), and cross-subject strategy adaptations (Section 5.3). These profiles provide students and educators with insights to make informed decisions about learning strategies.
Subject-Specific Cognitive Dimension. Our analysis (Section 5.3) reveals distinct cognitive patterns across subjects. Mathematics learning requires systematic logical derivation and involves longer reflection times for conceptual understanding. Physics benefits from graphical reasoning approaches combined with logical thinking, while chemistry favors efficient problem-solving strategies with concise solution methods. With data from a larger sample across multiple institutions and regions, these findings could be further validated to inform textbook development for enhanced thinking skill guidance and examination design with better-aligned question types and time allocations.
Limitations and Future Work. Our research also has several limitations, with opportunities for future work.
Dataset. The dataset in Section 3 used in this study is limited in both size and diversity, as it focuses only on science subjects (Physics, Chemistry, and Mathematics) from a single class of students. This narrow scope may not fully reflect the diversity of student populations across different academic streams, socioeconomic backgrounds, and learning needs. Future research could expand the analysis to encompass a wider variety of subjects, student demographics, and long-term studies to track individual performance trajectories over time.
MLLMs Integration. While CogChain effectively analyzes students' problem-solving patterns, current MLLMs primarily rely on surface-level pattern matching, which limits their ability to grasp deep reasoning and detect conceptual errors [21, 22]. Challenges remain in comprehending complex reasoning and generating adaptive feedback. Future research will focus on enhancing CogChain's capacity to model intricate cognitive processes, integrating more refined learning behavior databases, and advancing personalized AI-driven support [23].

Acknowledgments. This paper is supported by NSF of China (62402409), Guangdong provincial project 2023CX10X008, Guangzhou Municipality Big Data Intelligence Key Lab (2023A03J0012), Guangdong Basic and Applied Basic Research Foundation (2023A1515110545), Guangzhou Basic and Applied Basic Research Foundation (2025A04J3935), and Guangzhou-HKUST(GZ) Joint Funding Program (2025A03J3714).

References

1. Akben, N.: Effects of the Problem-Posing Approach on Students' Problem Solving Skills and Metacognitive Awareness in Science Education. Research in Science Education **50**(3), 1143–1165 (2020)
2. Arnold, I.J.: Do examinations influence student evaluations? International Journal of Educational Research **48**(4), 215–224 (2009)
3. Bhutoria, A.: Personalized education and Artificial Intelligence in the United States, China, and India: A systematic review using a Human-In-The-Loop model. Computers and Education: Artificial Intelligence **3**(100068), 1–18 (2022)
4. Bridgwood, M.: Guidelines for communication and engineering problem solving at the basic level. IEEE Transactions on Professional Communication **42**(3), 156–165 (1999)
5. Chaudhury, R., Chilana, P.K.: Designing Visual and Interactive Self-Monitoring Interventions to Facilitate Learning: Insights From Informal Learners and Experts. IEEE Transactions on Visualization and Computer Graphics **31**(2), 1542–1556 (2025)
6. Cho, Y.I., Altayeva, A.B.: Android-Based E-Board Smart Education Platform Using Digital Pen and Dot Pattern. The International Journal of Fuzzy Logic and Intelligent Systems **15**(4), 260–267 (2015)
7. Cirillo, M., Hummer, J.: Competencies and behaviors observed when students solve geometry proof problems: an interview study with smartpen technology. ZDM – Mathematics Education **53**(4), 861–875 (2021)
8. Docktor, J.L., Dornfeld, J., Frodermann, E., Heller, K., Hsu, L., Jackson, K.A., Mason, A., Ryan, Q.X., Yang, J.: Assessing student written problem solutions: A problem-solving rubric with application to introductory physics. Physical Review Physics Education Research **12**(1), 1–18 (2016)
9. Geraldi, J., Lechter, T.: Gantt charts revisited: A critical analysis of its roots and implications to the management of projects today. International Journal of Managing Projects in Business **5**(4), 578–594 (2012)
10. Guo, P., Shao, J., Ji, J., Pan, L., Pu, H., Pu, R.: Application status and prospect of dot matrix digital pen. In: Yu, Z., Patnaik, S., Wang, J., Dey, N. (eds.) Advancements in Mechatronics and Intelligent Robotics. pp. 545–549. Springer (2021)
11. Guruloo, T.N.M., Osman, K.: Beyond Pen and Paper: Understanding the Landscape of Digital Assessment in Stem Higher Education. In: Lecture Notes in Educational Technology, vol. Part F3676, pp. 177–197. Springer (2024)
12. Han, A., Zhou, X., Cai, Z., Han, S., Ko, R., Corrigan, S., Peppler, K.A.: Teachers, Parents, and Students' perspectives on Integrating Generative AI into Elementary Literacy Education. In: Proceedings of the CHI Conference on Human Factors in Computing Systems. pp. 1–17. ACM (2024)
13. Islam, M.S., Doumbouya, M.K.B., Manning, C.D., Piech, C.: Handwritten Code Recognition for Pen-and-Paper CS Education. In: Proceedings of the Eleventh ACM Conference on Learning @ Scale. pp. 200–210. ACM (2024)
14. Jovanovic, M., Vukicevic, M., Milovanovic, M., Minovic, M.: Using data mining on student behavior and cognitive style data for improving e-learning systems: a case study. International Journal of Computational Intelligence Systems **5**(3), 597 (2012)
15. Labahn, G., Lank, E., MacLean, S., Marzouk, M., Tausky, D.: MathBrush: A System for Doing Math on Pen-Based Devices. In: 2008 The Eighth IAPR International Workshop on Document Analysis Systems. pp. 599–606. IEEE (2008)

16. Li, B., Luo, Y., Chai, C., Li, G., Tang, N.: The dawn of natural language to sql: Are we fully ready? Proceedings of the VLDB Endowment **17**(11), 3318–3331 (Jul 2024)
17. Li, W., Hilton, E., Hammond, T., Linsey, J.: Persketchtivity: An Intelligent Pen-Based Online Education Platform for Sketching Instruction. In: Electronic Visualisation and the Arts. pp. 1–9 (2016)
18. Liang, Z., Sha, L., Tsai, Y.S., Gašević, D., Chen, G.: Towards the Automated Generation of Readily Applicable Personalised Feedback in Education. In: Proceedings of the 25th International Conference on Artificial Intelligence in Education, AIED 2024, pp. 75–88. Springer Nature Switzerland (2024)
19. Liu, J., Huang, Z., Xiao, T., Sha, J., Wu, J., Liu, Q., Wang, S., Chen, E.: SocraticLM: Exploring socratic personalized teaching with large language models. In: The 38th Annual Conference on Neural Information Processing Systems (2024)
20. Liu, X., Shen, S., Li, B., Ma, P., Jiang, R., Zhang, Y., Fan, J., Li, G., Tang, N., Luo, Y.: A survey of NL2SQL with large language models: Where are we, and where are we going? (2025), https://arxiv.org/abs/2408.05109
21. Liu, X., Shen, S., Li, B., Tang, N., Luo, Y.: Nl2sql-bugs: A benchmark for detecting semantic errors in nl2sql translation. arXiv preprint arXiv:2503.11984 (2025)
22. Luo, T., Huang, C., Shen, L., Li, B., Shen, S., Zeng, W., Tang, N., Luo, Y.: nvbench 2.0: A benchmark for natural language to visualization under ambiguity (2025), https://arxiv.org/abs/2503.12880
23. Shen, L., Li, H., Wang, Y., Xie, X., Qu, H.: Prompting Generative AI with Interaction-Augmented Instructions. In: Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, CHI EA '25. pp. 1–9. ACM (2025). https://doi.org/10.1145/3706599.3720080
24. Shen, S., Lu, S., Shen, L., Sheng, Z., Tang, N., Luo, Y.: Ask humans or ai? exploring their roles in visualization troubleshooting. arXiv preprint arXiv:2412.07673 (2024)
25. Siddiqui, A.T., Muntjir, M.: An Approach to Smart Study using Pen and Paper Learning. International Journal of Emerging Technologies in Learning **12**(05), 117 (2017)
26. Tang, N., Yang, C., Fan, J., Cao, L., Luo, Y., Halevy, A.Y.: Verifai: Verified generative AI. In: CIDR. www.cidrdb.org (2024)
27. Van Mechelen, M., Smith, R.C., Schaper, M.M., Tamashiro, M., Bilstrup, K.E., Lunding, M., Graves Petersen, M., Sejer Iversen, O.: Emerging Technologies in K–12 Education: A Future HCI Research Agenda. ACM Transactions on Computer-Human Interaction **30**(3), 1–40 (2023)
28. Walny, J., Lee, B., Johns, P., Henry Riche, N., Carpendale, S.: Understanding pen and touch interaction for data exploration on interactive whiteboards. IEEE Transactions on Visualization and Computer Graphics **18**(12), 2779–2788 (2012)
29. Wang, J.S., Chuang, F.C.: An Accelerometer-Based Digital Pen With a Trajectory Recognition Algorithm for Handwritten Digit and Gesture Recognition. IEEE Transactions on Industrial Electronics **59**(7), 2998–3007 (2012)
30. Wang, Y., Chiew, V.: On the cognitive process of human problem solving. Cognitive Systems Research **11**(1), 81–92 (2010)
31. Yan, Y.: Analysis of Teachers' and Students' Technology Application Behaviors in a Smart Paper-Pen Environment. In: Proceedings of the 2023 6th International Conference on Educational Technology Management. pp. 42–48. ACM (2023)
32. Yang, Q., Zhan, Z., Du, X.: A Learning Engagement Analysis Model for Assignments based on Dot Matrix Pen Technology. In: Proceedings of International Conference on Education and Multimedia Technology. pp. 90–96. ACM (2022)