J Vis (2024) 27:235–254
https://doi.org/10.1007/s12650-024-00956-4

REGULAR PAPER

Zhiguang Zhou · Lihong Cai · Jiayi Guo · Yong Zhang · Baofeng Chang · Ting Xu · Yongheng Wang

ExeVis: concept-based visualization of exercises in online learning

Received: 5 November 2023 / Accepted: 23 December 2023 / Published online: 14 March 2024

Abstract In recent years, online learning has gained popularity and proven to be an effective way of education. Numerous studies have analyzed teaching materials and learning behaviors. However, most of the existing studies ignore the relationships between learning concepts and exercises, which can convey teaching performance and student behaviors. Presenting the relationships between concepts and exercises in online learning not only can help educators explore the distribution of exercises and concepts to consolidate knowledge but also can provide intuitive feedback on student behavior in online courses, which can enhance the teaching strategy. In this work, we extract learning concepts from exercises, establish logical relationships between concepts and exercises, and construct the hierarchical structures of concepts via both automatic models and semi-automatic models. To help users analyze and evaluate concepts and exercises effectively and intuitively, we design and implement a visual analysis prototype system, named ExeVis, integrating multiple interactive visualization graphs. ExeVis is equipped with multiple interactive and intuitive visualization charts including a control view to select and display basic information, an overview with hierarchical structures to present the distribution and mastery of concepts and exercises, a correlation view to reveal relationships between exercises, and a performance view to show individual capability. Case studies with real data and expert interviews demonstrate the usefulness and effectiveness of ExeVis in providing educators with valuable insights into the appropriateness of exercises and enabling them to adjust their teaching methods.

Z. Zhou · L. Cai · J. Guo · Y. Zhang
Intelligent Big Data Visualization Lab, Hangzhou Dianzi University, Hangzhou 310000, China
E-mail: zhgzhou@hdu.edu.cn

L. Cai
E-mail: cailihong@hdu.edu.cn

J. Guo
E-mail: guojoeeeeeee@gmail.com

Y. Zhang
E-mail: zhangyong@hdu.edu.cn

B. Chang (✉)
Zhejiang Airport Innovation Institute, Hangzhou 310000, China
E-mail: baofeng.chang@foxmail.com

T. Xu
Department of Stomatology, First Affiliated Hospital, Zhejiang University, Hangzhou 310000, China
E-mail: xut@zju.edu.cn

Y. Wang (✉)
Research Center for Data Mining and Knowledge Discovery, Zhejiang Lab, Hangzhou 311121, China
E-mail: wangyh@zhejianglab.com

Keywords Online learning · Education analysis · Exercise evaluation

1 Introduction

Online learning has emerged and becomes one of the most popular ways to help people access knowledge and skills in recent years. With the increasing number and diversity of online learning, numerous studies have been conducted to analyze courses through various metadata, such as videos, slides, syllabus, and exercises (Means et al. 2009; Ferguson 2012; Zhang et al. 2022). Specifically, many current works transform knowledge down to micro-level units and classify them into different groups to improve the efficiency of concept analysis (Ferguson and Clow 2015; Nesbit and Adesope 2006). Furthermore, many works also pay attention to evaluating learner behaviors with learning materials, such as forum posts, video-watching behavior, course notes, and dropout rates (Ferguson and Clow 2015; Zheng et al. 2015; de Barba et al. 2020). Such works not only assist educators in understanding students' learning processes but also facilitate the adaptation and customization of courses (Merceron and Yacef 2005; Romero and Ventura 2010). However, most existing online learning analysis works only regard exercises as a means of performance measurement for validation and accreditation, overlooking the value of analyzing exercises in depth to enhance the learning process (Mayer 2005; Chen et al. 2015).

Exercise, derived from a subject syllabus, naturally exposes learning concepts to students (Bransford et al. 2000). Thus, exercise is an indispensable teaching tool for educators, aiding students in reinforcing the knowledge that is acquired in class (Anderson 1996; Mayer 2005). Moreover, the exercise plays a critical role in providing feedback for educators/teachers (Hattie and Timperley 2007). Currently, many online learning platforms record web log data including both basic information about exercises and learner behavior with exercise (e.g., exercise accuracy, the number of attempts, open time, and due time), which are the basic data of many visual analysis system that analyzes and visualizes online learning exercise data (Chen et al. 2018; Emmons et al. 2017; Mazza and Dimitrova 2007). Nevertheless, current systems primarily concentrate on explicit students' behaviors through statistical evaluation techniques, yet they lack a thorough exploration of the intricacies within exercises, consequently failing to reveal potential underlying patterns and insights.

To bridge this gap, it is crucial to approach exercises from a conceptual perspective. Analyzing the relationship between exercises and underlying learning concepts can yield valuable insights to enhance teaching methods. For instance, in online programming courses, a series of coding tasks are presented. By analyzing score rates and attempts, we can gain insights into students' behavior and identify concepts they may struggle with.

Moreover, to further identify issues and requirements, we conducted interviews with education experts who emphasized customized exercises with specific focuses and varied formats (Pellegrino 2002; Wiliam 2011). They highlighted the importance of monitoring learner behavior, assessing comprehension levels, grouping students based on problem-solving abilities, and identifying outliers. By integrating these insights with a thorough literature review, we have gained an improved understanding of the problems, which can be summarized into three major challenges (CH1–CH3). Firstly, comprehensive extraction and sorting of concepts from exercises pose difficulties due to the mixture of explicit and implicit concepts (CH1). Secondly, externalizing and visualizing the knowledge structure related to exercises and student behaviors across different levels present a crucial challenge (CH2) (Pan et al. 2017). Finally, establishing logical relationships among concepts and exercises is challenging due to their complex and sometimes latent associations (CH3).

To tackle these challenges, we propose a semi-automatic method for extracting learning concepts from exercises, enabling the analysis of relationships between concepts and exercises. Additionally, we propose ExeVis, a comprehensive visualization system that provides exercise content analysis and intuitive feedback on student performance to educators (e.g., teachers, course instructors, and education analysts). ExeVis displays hierarchical structures of concepts and exercises to analyze the appropriateness of content and distribution. Meanwhile, it offers an overview of multi-scale learner behavior for concepts, exercises, and their dependencies, allowing users to identify general learner patterns and trends. Furthermore, we design a network diagram to reveal the implicit relationship between exercises and concepts. For students, we visualize individual performance in a multidimensional manner, facilitating the identification of similarities and differences. The major contributions of this paper are as follows:

- A analysis pipeline consists of a semi-automatic concept extraction method and a visual analysis component, enabling knowledge exploration in various exercise analysis scenarios.
- A visual analytics system with interactive diagrams facilitates users in intuitively exploring relationships among exercises, concepts, and learner behavior across multiple dimensions.
- Insights derived from authentic online exercise data empower educators to adapt exercise settings and enhance teaching approaches.

2 Related work

2.1 Online learning analysis

In recent years, online learning has emerged as a popular educational delivery method (Means et al. 2009; Kui et al. 2022). To facilitate student learning and teacher analysis, numerous studies have analyzed different metadata in online courses, including videos (Huang et al. 2017; Liu et al. 2018; Schwab et al. 2016), slides (Chen et al. 2010; Wei and Chou 2020), syllabus (Huang et al. 2019; Dai et al. 2016), and exercises (Chu et al. 2006; Means et al. 2009). For instance, Chen et al. explored the impact of three commonly used video lecture styles on learners' sustained attention, emotion, cognitive load, and learning performance. Moreover, researchers have delved deeper by integrating multiple types of course content to conduct comprehensive investigations. For example, Koedinger et al. (2015) categorized course materials into passive/declarative information and active/interactive activities, investigating the learning benefits associated with each.

Due to the benefits of concept mapping in subject areas (Nesbit and Adesope 2006), plenty of studies focus on evaluating concepts and their interrelationships extracted from the multimodel cues (Zhao et al. 2017; Zhang et al. 2022). For example, Chen et al. (2010) proposed several approaches to extract key terms from spoken course lectures, while Zhang et al. (2019) enabled users to manually create relationships between concepts using frequency and timestamped representations of keywords. Moreover, Liu et al. (2018) and Schwab et al. (2016) developed concept maps to help learners quickly associate concepts with videos. Cooper et al. (2018) combined both video content and sequential inter-topic relationships mined from the course syllabus and provided an interactive visual exploration of recommendations.

However, exercise evaluation, crucial for reinforcing student learning and providing feedback to teachers, has received limited attention. And the majority of online learning methods only rely on exercise grades to measure performance and validate the efficacy of other materials (Chen et al. 2015), neglecting the wealth of information contained in exercises.

2.2 Learner behavior visualization

Monitoring students' learning behavior is crucial for effective online learning (Helic et al. 2000). However, the data obtained from online learning platforms are often complex, massive, noisy, and difficult to comprehend (Zhang et al. 2022). To address this issue, many studies have employed visualizations to intuitively demonstrate their analysis (Wang et al. 2023).

Some studies have used basic visualizations like scatter charts and line charts. For example, Mazza and Dimitrova (2007) used these charts to identify trends and individuals needing attention, while Huang et al. (2016) employed stacked bar charts and river maps to visualize course enrollment, engagement, and performance. Huang et al. (2017) focused on mining learning concepts from learners' behavior data and subtitles and represented them using a keyword cloud. Moreover, Mubarak et al. (2021) visualized the video-clickstream data generated by learners' interaction with course videos as graph networks. Recently, more comprehensive visualization systems have emerged. For instance, Chen et al. (2015) presented a visualization system focusing on interaction peaks, with intricate glyph designs such as glyphs, flow maps, and parallel coordinates to display multiple attributes of a peak. Chen et al. (2016) developed a visualization system that integrates four linked visual designs to help analysts identify learning patterns related to dropout behavior at multiple scales and reveal detailed learning activities of learners.

In summary, while past studies have commonly utilized multi-panel views and interactive techniques like sorting, filtering, and clustering to demonstrate learning behavior, our method offers a conceptual analysis of exercises' behavior. We go beyond traditional visualizations and propose an analytics system that enables users to gain a multifaceted understanding of learning behaviors. This not only provides a comprehensive and nuanced comprehension of exercises and concepts but also offers valuable insights into group dynamics and individual behaviors.

2.3 Exercise analysis

Several researchers have explored exercise evaluation in various ways. For example, Mazza and Dimitrova (2007) created a matrix to track students' attempts at questions related to different concepts. Grover et al. (2014) and Tomkins et al. (2016) visualized grade distributions for each exercise. Moreover, Stephens-Martinez et al. (2014) used stacked bar charts to investigate attempt distribution. Emmons et al. (2017) analyzed exercise performance at macro- and micro-levels, comparing student submissions to identify problematic areas. Recently, Arruarte et al. (2021) developed TEA, a visual analytics tool that analyzes and visualizes student performance on test-based exercises, providing indicators for exercise validity and quality improvement when they do not meet teachers' expectations.

In conclusion, previous studies have focused on using exercise behavior to provide feedback on content revision and teaching methods. However, they overlook the underlying knowledge that exercises are designed to reinforce and consolidate (Bransford et al. 2000; Han et al. 2013; Anderson 1996; Mayer 2005; Brown et al. 2014; Dunlosky et al. 2013). Additionally, these studies rely on statistical evaluation techniques that have limitations in capturing complex relationships and identifying patterns and trends in learner behavior and performance (Mayer 2005; Dunlosky et al. 2013). Without such analysis, it becomes challenging to fully comprehend the effectiveness of exercises in promoting learning outcomes. As far as we know, there is currently no tool available that can effectively visualize exercises across various levels of content and behavior.

3 Task and system workflow

3.1 Data description

The exercises utilized in this study were designed by actual educators for a C Language Programming course and were administered to a class of 40 freshmen majoring in digital media technology. Most of them lacked practical experience and a deep understanding of knowledge and skills in areas such as programming. According to the course curriculum and student needs, instructors created weekly, unit and summary exercises. The exercises and corresponding student behaviors were logged in PTA, an online experiment-assisted teaching platform. Due to the complex and multidimensional log data (Mazza and Dimitrova 2007), we preprocessed it by filtering out irrelevant characters and selected key measurements like submission status, attempts, and time. After reviewing the exercises, we identified 236 questions from the first to the twelfth week, covering a range of question types such as multiple choice, fill-in-the-blank, and true/false. Specifically, the exercises included coding tasks that are critical for programming language teaching.

3.2 Task analysis

Before summarizing tasks, we conducted several rounds of interviews with four domain experts (E1–E4) to gather feedback on the exercises from various perspectives. E1, as an administrator of courses in a university, is responsible for curriculum development and implementation and ensures the student obtains a high-quality education. E2 is a course instructor for C Language Programming, with rich experience in education in both online and offline courses. Additionally, she participates in the creation of textbooks and other instructional materials. E3 is an expert in teaching and has been offering online lecture videos on MOOCs for over ten years. She has a deep understanding of pedagogy and how to effectively engage and support students in an online learning environment. E4 is an educational analyst for online learning, with a focus on understanding and improving student outcomes. He analyzes student data and provides feedback about improvements in the curriculum or delivery methods. Additionally, we conducted a literature review to further refined the tasks.

The tasks (T1–T4) of this work are introduced as follows:

T1. How to comprehensively extract concepts? When designing exercises, educators often include multiple types of problems that combine related concepts to reinforce learning. However, most log data only record major concepts related to the question, potentially leaving out other important information.

Additionally, some underlying concepts may be difficult for current entity extraction models to identify. Therefore, addressing these challenges requires models that are knowledgeable in multiple fields to extract all relevant concepts and determine their relevance to the exercises, ensuring that information is organized in an orderly and accessible manner (Tukey et al. 1977; Schwab et al. 2016; Yang et al. 2024).

T2. What feedback can be obtained from learner behavior? The variety of log data provides multidimensional insights into learner behavior. For instance, accuracy is the most common metric used to measure students' understanding of the concepts covered in the exercises. Specifically, in contrast to traditional classroom teaching, online learning allows for multiple attempts, which can indicate the level of effort required for mastery and help teachers identify areas where students struggle the most. Drawing on students' multidimensional performance can provide valuable insights for educators to gain a comprehensive understanding of learner behavior and concepts at different levels. This can reveal individual strengths and weaknesses, as well as patterns and trends among similar groups of students. Besides, this information can help educators identify weak areas in specific knowledge points, pinpoint poorly-performing units, and gain a better understanding of the overall mastery of the course.

T3. What are the logical relationships among exercises and concepts? Analyzing the associations between exercises and concepts is crucial for effective teaching and optimizing instructional practices. These relationships can be categorized into two types: explicit and implicit. Explicit relationships are based on the concepts extracted from exercises, which naturally connect exercises containing the same or related concepts. Through these relationships, educators can analyze the mastery of concepts and find dependencies and prerequisite knowledge required for exercises. However, some concepts have underlying connections that are not found during extraction. By analyzing the order of exercises and student performance across multiple exercises, implicit relationships can be revealed. These relationships can help educators sequence exercises in a way that is pedagogically sound.

T4. How to evaluate the appropriateness of exercise design? To enhance students' mastery of the course, exercise design should align with learning goals. Previous research and expert interviews have demonstrated that content distribution and exercise order are crucial factors. Thus, careful consideration should be given to the appropriate number, range, and order of exercises needed to reinforce learning objectives. Additionally, feedback from student behavior can determine if exercise settings are appropriately challenging and provide opportunities for students to apply and extend their understanding. By analyzing both distribution and feedback, educators can identify areas of difficulty and adjust their exercise design accordingly.

3.3 System workflow

To address the tasks mentioned above, we present ExeVis, a visual analysis prototype system that provides educators with valuable insights into exercise distribution and intuitive feedback on student performance. As demonstrated in Fig. 2, the system pipeline starts with extracting concepts from exercises, followed by association rule mining to discover fundamental implicit relationships between exercises. The multidimensional behavior of each exercise, concept, and individual is then calculated. Finally, a hierarchical structure is defined to present the distribution and behavior of multilevel concepts and exercises. Our visualization system features four coordinated views: a control view, an overview, a correlation view, and a performance view, as illustrated in Fig. 1. Furthermore, we have incorporated a rich set of interactions that enable users to explore regions of interest flexibly and effectively.

4 Approaches

4.1 Concept extraction

The concepts and relationships contained in exercises are complex and sometimes underlying, so it is necessary to use models for extraction. To establish a clear hierarchy of concepts, we leverage the course syllabus as prior knowledge and create an initial structure. However, syllabi may not provide the comprehensive and detailed information we need as they primarily focus on important teaching topics (Huang et al. 2019). To address this issue, we utilize entity extraction algorithms, including the NLTK toolkit (Loper and Bird 2002), as well as models based on BiLSTM-CRF (Zhong and Tang 2020) and BERT (Devlin et al. 2018). While these methods perform well in extracting obvious concepts, they struggle

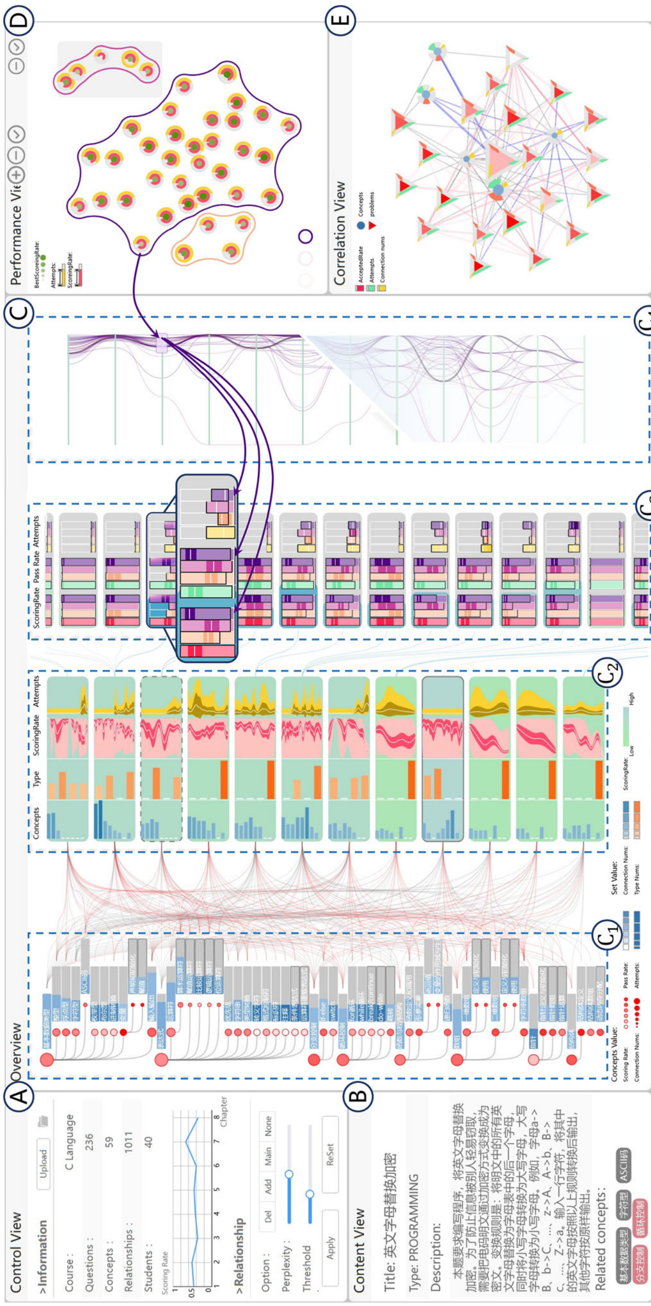

Fig. 1 ExeVis system. **A** Provides basic information display and selection, and **B** visualizes question content with related concepts. Overview **C** consists of the hierarchical structure of concepts (C_1), exercise and problem distribution (C_2 and C_3), and overall/individual behavior (C_4). Performance view **D** shows individual performance, while correlation view **E** reveals exercise relationships

to identify implicit concepts and their interrelationships due to limited knowledge across various course fields. Additionally, they require substantial amounts of data and are time-consuming. Therefore, we require a model that possesses both comprehensive knowledge of various fields and high processing speed.

Existing studies have validated the satisfactory performance of GPT series models in text extraction tasks (Wei et al. 2023; Li et al. 2023a, b; Han et al. 2023), demonstrating their accurate capture of entity boundaries and contextual information for high-quality extraction. To evaluate GPT-3.5's ability in exercise concept extraction, we conducted a comparative experiment with course instructors. The results showed that GPT-3.5 performs at a comparable level of quality to human experts, offering the advantages of speed and occasionally providing more detailed and comprehensive output.

Given the limitations of existing models in concept extraction due to data availability, we employed a semi-automatic approach. Firstly, we utilized the OpenAI API (OpenAI 2023) to access GPT-3.5 for concept extraction. To ensure consistent and coherent responses aligned with the chosen course, we endowed GPT-3.5 with the persona of an expert in the specific course, providing exercise content and course/chapter details. The output was constrained to match the chosen course's expertise and stored in a database for future use. Subsequently, we utilized GPT-3.5 to merge the extracted concepts with those already recorded in the syllabus to create the concept hierarchy. However, challenges such as concept name variations, misplacements, and omissions affect the reliability of the concept hierarchy. To mitigate these issues, we employed multiple rounds of dialog-based evaluation to assess the quality of the generated results, ensuring greater stability. Furthermore, we performed concept merging based on their similarity and categorized the relationships between concepts as primary or subordinate. Lastly, we ensure the accuracy of the results through manual verification.

4.2 Learner behavior analysis

The log records a large amount of data, providing multiple perspectives for analyzing learners' behavior. In order to accurately reflect students' performance, we have selected score rate as the main measurement. Besides, compared to traditional offline education, online learning platforms provide students with opportunities to attempt several times. As demonstrated in many researches (Ahadi et al. 2016; Davis et al. 2020), multiple attempts allow students to learn from their mistakes and lead to better learning outcomes. In addition, it encourages students to actively engage with the exercise and can improve their understanding of the subject matter. Therefore, we consider attempts as a means of measurement in our analysis.

We have a set of exercises E with k exercises, each containing a set of questions $E = q_{11}, q_{12}, \ldots, q_{ij}, \ldots, q_{kl}$. Here, q_{ij} refers to the j-th question in the ith exercise, and l represents the number of questions in kth exercises. We then calculate the average scoring rate for each question. Furthermore, based on the record, we determine the pass rate which means the proportion of students who answer the question correctly before the due time.

Based on these data, we further evaluate the behavior of each student, concept, and exercise. We represent the student's performance on each question as a feature vector, with each question corresponding to a specific dimension. The value in each dimension represents the student's scoring rate, calculated by averaging the score per attempt. This high-dimensional vector (Zhou et al. 2020) provides valuable insights into the student's characteristics. To reduce the dimensionality of the data, we utilize t-SNE (Van der Maaten and Hinton 2008) after comparing it with other dimension reduction algorithms like MDS (Zimet et al. 1988), PCA (Abdi and Williams 2010), and UMAP (Becht et al. 2019). By integrating these data with previously extracted relationships, we can evaluate the behavior of concepts. For exercises where the concept is the primary focus, we assign a weight of 0.8, while a weight of 0.2 is assigned to other exercises. This weighting scheme enables a more accurate assessment of concept behavior.

4.3 Exercise relationship extraction

After extracting concepts from exercises, we establish the fundamental relationships between the structure of concepts and exercises, which can reveal the explicit interrelationships between exercises. By utilizing the concepts contained in exercises, it becomes easier to locate other exercises that share the same concept. Additionally, the distribution of exercises that regard a particular concept as the key knowledge, as well as those that only contain it, can be identified. These relationships enable us to gain valuable insights into distinguishing between similar exercises and identifying dependencies exercises.

Apart from the explicit relationships found through concepts, some exercises have underlying associations which can be revealed by analyzing the learner's behavior. For example, some students may perform better if two exercises are in a special order. To uncover these implicit relationships between exercises, we utilize association rule mining, a widely-used technique in academic research for improving qualitative teaching aspects (Vranic et al. 2007; Romero and Ventura 2010).

First, we define an association rule $X \Rightarrow Y$ and calculate their support, where X and Y are individual questions. The function *support*(X) represents the proportion of students who attempted the question X but still failed before the deadline, and is calculated as follows:

$$support(X) = \frac{count(X_{fail})}{num} \tag{1}$$

where $count(X_{fail})$ is the number of students who attempted question X and answered incorrectly, and *num* represent the number of students.

Next, we calculate the confidence between questions, which indicates the percentage of students who answered question Y incorrectly given that they failed question X. Confidence can be calculated through the following formula:

$$confidence(X \Rightarrow Y) = \frac{support(X \cup Y)}{support(X)} \tag{2}$$

where $X \cup Y$ refers to students who attempted both questions X and Y.

Afterward, we employ lift to filter out valuable question pairs. Lift determines whether answering question Y would be easier or harder when the student has already mastered question X and can be calculated using the formula:

$$lift(X \Rightarrow Y) = \frac{confidence(X \Rightarrow Y)}{support(Y)} \tag{3}$$

Finally, we select question pairs whose lift is greater than 1 and whose confidence exceeds a certain threshold. By performing these calculations, we can extract valuable question pairs and help educators better understand the underlying relationships between questions.

5 ExeVis

5.1 Design requirements

The primary design requirements (R1–R3) are summarized as follows:

R1. Provide intuitive representations of content and behavior at different levels The system should offer visual representations of exercise and concept distribution and behavior at various levels, such as individual student performance, exercise behavior, and concept distribution. Visual encodings, like color-coded schemes and hierarchical structures, help educators identify patterns and relationships. Clear explanations of the visualizations, including instructions on data interpretation and system usage, support informed teaching decisions.

R2. Enable quick identification of associations and differences between exercises and concepts To facilitate the understanding of the relationships between exercises and concepts, the system should provide clear and distinguishable visual representations of primary and secondary contacts, as well as two types of connections between exercises. Specifically, the correlation coefficients need to be accurately measured and mapped to appropriate visual cues to convey differences. Moreover, multiple layers should be available to present the multidimensional learning behavior for different levels of concepts, exercises, and individuals. Additionally, the system should allow for easy comparison of behaviors during and across different units.

R3. Allow for interactive filtering with immediate feedback To support user intervention, the system should offer familiar interactions, such as selecting and filtering specific exercises and concepts, dragging and dropping for relationships, and lassoing circles for selecting a user-defined group. According to the performance of those interactions, the system should provide immediate feedback, such as fading, highlights, and tooltips, to show the resulting changes in the visualization. For instance, if educators select a particular group of students in the performance view, the system should display a corresponding behavior distribution to indicate how the selected group of learners performs.

5.2 Visual design

To address the tasks, requirements, and user-centered design principles (Munzner 2009), we design a visual analysis system, named ExeVis, to help users analyze the online learning exercise data. ExeVis integrates many visualization diagrams including a control view (Fig. 1A) to display basic information and facilitate configuration changes, a content view (Fig. 1B) to present questions content and their corresponding concepts, an overview (Fig. 1C) to demonstrate the interconnectedness of concepts and exercises, a performance view (Fig. 1D) to show individual behavior, and a correlation view (Fig. 1E) to reveal relationships between exercises. These views incorporate common metrics such as scoring rate and number of attempts, distinguished by different colors and shading to represent behavior. In the control interface, users can access basic course information and adjust hyperparameters (Gove et al. 2022) by using sliders to fine-tune the t-SNE and layout results. The content view displays question information and related concepts. Detailed visual designs used in other views are presented in the following sections.

5.2.1 Overview

The overview plays a central role as the entry point for users, providing a hierarchical structure of concepts and exercises. As shown in Fig. 1C, it consists of four parts: C_1 displays the concept structure, C_2 illustrates the distributions of exercises, C_3 introduces corresponding questions, and C_4 showcases group behavior.

Based on their hierarchical relationships (T1), we utilize a vertical tree structure (Fig. 1C_1) where position exercises and questions are positioned based on their order of occurrence (R2). The chapter name is displayed at the first level, followed by sequential nodes downward. Specifically, concepts are represented by signal points capturing their position and relationships, along with content labels providing detailed information such as concept names. This approach optimizes space in the visual interface and enables more concepts to be displayed simultaneously. To enhance comprehension, we utilize color coding for concepts, exercises, and questions (R1). Concept points and exercise/question backgrounds reflect the average scoring rate, while content labels indicate the pass rate. The shading of colors corresponds to the scoring rate, with darker colors representing higher rates. In addition, the size of the outer radius indicates the number of related concepts, with larger sizes indicating more connections (R1). Besides, content labels use length to represent attempts, while questions employ length to indicate the number of related concepts, with the length directly corresponding to the quantity (R1). To differentiate between syllabus concepts and newly added ones, newly added concepts have a solid border, while the initial concepts from the syllabus are borderless. Similarly, exercise types are distinguished by their border style, with weekly exercises having no border, unit exercises having a dashed border, and summarized exercises having a solid border. As for the relationship, color is employed to differentiate and highlight the main and subordinate relationships due to their significance (R2).

To visualize the diverse distribution of exercises, we utilize four parallel charts (R1). The first two bar charts display the distribution of chapters and question types, with the length and color corresponding to the quantity. The order in the first chart represents the chapter order, while the second chart represents true or false, multiple-choice, fill-in-the-blank, and programming tasks. For a detailed understanding of scoring rate, pass rate, and attempts, we designed stacked charts (Fig. 3a) that showcase the minimum and maximum scores/attempts, average scores/attempts, and variance for exercises and questions.

After selecting a specific exercise in Fig. 1C_2, the bar charts display the scoring rate, pass rates, and attempts for each question within the exercise. The upper and lower borders, as well as the varying shades of color in the regions, respectively, indicate the maximum and minimum values, upper and lower quartiles, while the middle white line represents the average. If student grouping is specified through the performance view, the respective information for each group will also be displayed on the right side of the corresponding bars and they are distinguished by colors corresponding to the performance view like Fig. 1.

Furthermore, we employ parallel coordinate plots (Firat et al. 2022) to visualize the performance trends of individual students across exercises (R1). Each axis represents the scoring rate, with higher rates indicated from left to right. Specifically, the bold lines represent the average values for all students. In the initial interface, each axis corresponds to a specific exercise set in Fig. 1C_2. Upon selecting a particular exercise group, the details of the questions are displayed in Fig. 1C_3. Each axis then corresponds to each question in Fig. 1C_3. When a student group is selected, the group behavior's line colors correspond to the group color shown in the performance view.

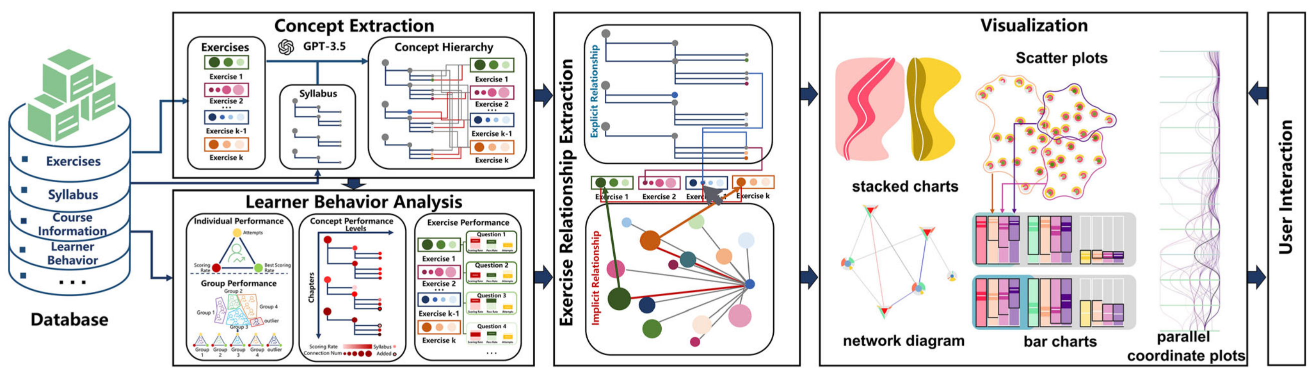

Fig. 2 The ExeVis system pipeline facilitates a systematic flow of concept extraction, learner behavior analysis, and exercise relationship extraction. Additionally, it provides interactive visualization designs to present exercise distribution and student performance

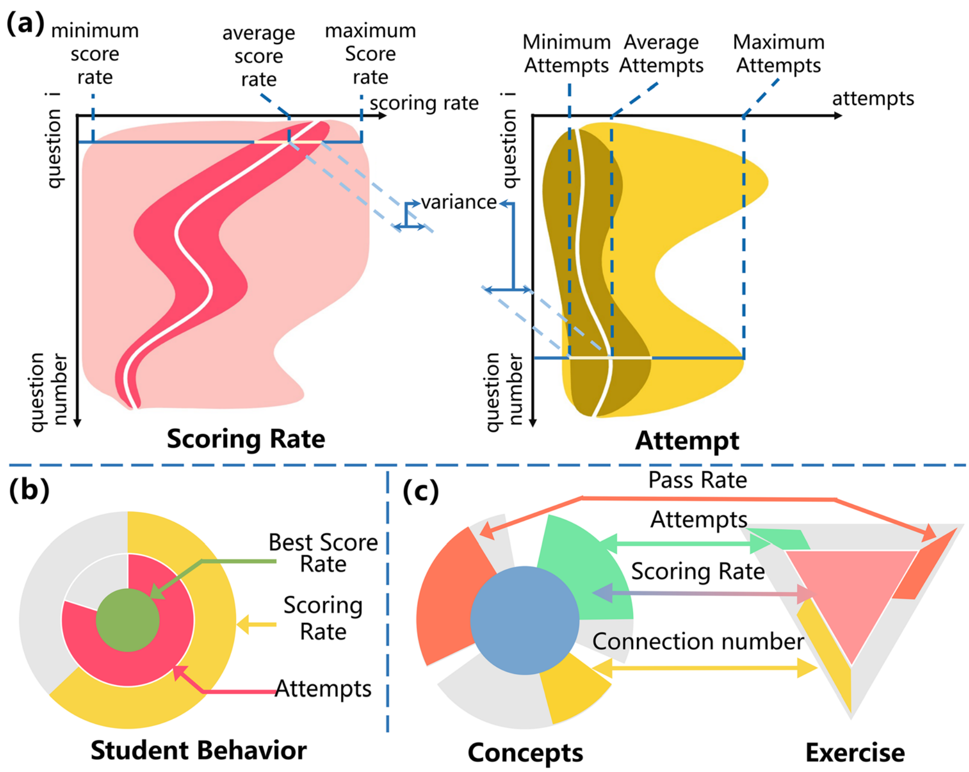

Fig. 3 **a** Stacked charts are utilized to visualize the distribution of scoring rates and attempts for all exercises. **b** Three key attributes of student behavior, namely the best score rate, scoring rate, and attempts. **c** The attributes of concepts and exercises, including scoring rate, pass rate, number of connections, and number of attempts

5.2.2 Performance view

To visually represent each student's multidimensional performance (R1), we have developed a glyph composed of a filled circle and two outer rings (Fig. 3b) to demonstrate student engagement. The central circle represents the best scoring rate, while the inner and outer rings represent the number of attempts and the scoring rate. Besides, the circle's background color maps to the student's best score, with darker shades indicating better performance. The outer ring is divided into sectors, with each sector's angle corresponding to the proportion of the student's behavior relative to their overall performance. This clockwise coding scheme enables clear interpretation and comparison of student performance. Moreover, as the glyphs become closer, it indicates that these students exhibit more similar patterns and trends.

5.2.3 Correlation view

To enhance the analysis of exercise relationships (T3), we utilize heterogeneous network diagrams (Zhou et al. 2022; Han et al. 2022; Peng et al. 2023) to visualize the connections between concepts and exercises. Besides, we employ shape encoding to represent them effectively (Fig. 3c). Exercises are represented by equilateral triangles, while concepts are depicted as circles. To incorporate other behaviors, we have introduced an outer ring divided into three segments. Each segment is color-coded to represent different behaviors, with the size of each color block determined by the proportion of specific behavior relative to the total behavior.

To enhance the understanding of different relationships between concepts and exercises, we introduce color differentiation to indicate the relationship between exercises and their main concepts. Additionally, we assign colors based on closeness derived from association rule mining. Moreover, only relationships exceeding a confidence threshold are displayed, offering meaningful insights. By incorporating these design choices, we provide users with a visual representation that intuitively conveys the relationships and their significance.

5.2.4 Interaction

ExeVis incorporates a range of intuitive interactions that help users analyze the exercise using the multiple visualizations discussed previously.

Interaction 1: interactive selection and customization Users can seamlessly navigate ExeVis and explore corresponding results using interactive selection techniques, including click, drag, and lasso selection. For instance, users can select specific concepts or exercises to view related questions, and this functionality extends to the bar charts depicting chapters and question types. The correlation view adjusts the center glyph to display relationships among selected items. If users encounter issues with concept-exercise relationships, they can modify visual attributes using the control view. By selecting the desired relationship type and dragging the corresponding association from the question to the correct concept, users can make changes that will be saved for future reference. Additionally, users can create self-defined groups and color-code the students belonging to these groups in the parallel coordinate plot, allowing them to compare the multidimensional behavior of each question among different groups.

Interaction 2: tooltip and highlighting To enhance effective data exploration and increase user engagement, ExeVis incorporates tooltips in various components. When users interact with elements like concepts and exercises in the overview or glyphs in the performance view or correlation view, tooltips appear, providing contextual information and explanations such as content, multidimensional behavior, and related concepts/exercises. In situations where clicking is not suitable, highlighting is employed to draw attention to specific data points or areas of interest. For instance, in the correlation view, clicking on a glyph may disrupt the identification of meaningful relationships. Therefore, we utilize highlighting techniques when hovering over exercises or concepts in the overview. This highlighting functionality is also implemented in the performance view and the parallel coordinate plot, enabling users to easily focus on relevant information.

6 Evaluation

To demonstrate the effectiveness and usefulness of ExeVis, a case study was conducted in collaboration with domain experts (E1–E4) during the task analysis process. The experts were initially introduced to the system to familiarize themselves with its features. They were then encouraged to freely explore the system, engaging in discussions and documenting their focal points, discoveries, and suggestions. Finally, we conducted in-depth interviews with the experts and administered questionnaires to other teachers. Through a combination of qualitative and quantitative analysis, their feedback was collected and evaluated to assess the strengths and weaknesses of our system.

6.1 Multidimensional distribution analysis

During the experts' free exploration, we observed their focus on specific distributions, which can be summarized as follows:

Distribution of concepts in exercises First of all, they examined the relationships between concepts and exercises present in the overview (Fig. 1C). They discovered that utilizing color differences helped them easily distinguish between main and subordinate relationships, despite their complexity. E2 and E3 then proceeded to browse the exercises sequentially, delving into the specific relationships of each question. E2 noticed some discrepancies between the syllabus and the hierarchical structure of concepts. "While designing the syllabus, I always ensure coverage of all essential topics, which is limited in depth at certain concepts. For instance, in the expression chapter, the syllabus only mentioned operators, but the border difference indicates the extraction source. I observed that the system provides more detailed information about operators, such as arithmetic operators, assignment operators, comparison operators, logical operators, and bit-wise operators. It does an excellent job of capturing and merging all the extracted concepts with the syllabus(T1)." Furthermore, drawing upon her teaching experience, E3 matched each question with the main and subordinate concepts. She remarked, "This visualization is convenient. When I click on an exercise or a question, the highlighting feature allows me to easily identify the related concept and the concepts contained in specific questions are also shown with its content. It enables me to quickly grasp the scope and focus of each exercise, facilitating the assessment of their appropriateness and guiding me in assigning more suitable exercises(T4)."

Following their exploration, the experts (E2 and E3) turned their attention to the first statistical charts within each exercise, which depicted the distribution of chapters. E2 observed that the chapter distribution in each exercise aligned well with their teaching plan. She emphasized the importance of aligning homework exercises with the ongoing curriculum and course pace, covering a range of topics taught in class to support student practice and comprehension. E2 found assessing alignment and comprehensiveness challenging in traditional approaches but acknowledged the visualization's effectiveness in evaluating exercise coherence and identifying exceptions. The visualization allowed her to easily evaluate the contribution of each exercise and detect potential issues such as premature introductions, gaps, or overlaps in course content coverage. She commented, "The first two chapters in C Language Programming establish the foundational knowledge for subsequent chapters. From this view, I also find this phenomenon, which demonstrates the appropriateness (T4)." During further exploration, E3 discovered an anomaly in the second exercise where questions related to later chapters, specifically the sixth and eighth chapters, were included. By clicking on the eighth chapter bar, she swiftly identified a question in the second exercise that pertained to a concept introduced much later in the curriculum. She remarked, "Thanks to the click interaction, I can easily identify questions that are related to the clicked chapter. This is very convenient." While analyzing the question content, E3 added, "Though students can complete the question after learning the first two chapters, it would be more beneficial for them to do so after studying the eighth chapter. This system does have the capability to comprehensively extract all the concepts from the questions(T1). "

Distribution of types and quantity After examining the concept distribution in the exercises, experts shifted their focus to evaluating the appropriateness of exercises designed for specific concepts. As this is a programming course for first-year students majoring in digital media technology, E1 highlighted the students' limited programming experience but emphasized their enthusiasm and eagerness to learn and improve their programming skills. Hence, the selection of exercise types, quantity, and question types is crucial.

After analyzing the variation in exercise borders, E2 highlighted the importance of click interactions to display all the questions within a specific exercise. She noted, "The visualization successfully differentiates between exercise types. Obviously, chapter exercises and summarized exercises prioritize key concepts. For instance, introductory chapters focusing on fundamental and straightforward concepts benefit from summarized exercises to assess students' mastery and skills. However, for a critical and challenging unit like the pointer chapter, unit exercises are necessary to evaluate students' comprehension, as the outcomes significantly impact their understanding of subsequent topics."

Afterward, experts delved into the first chart, where the length corresponds to the quantity, to analyze the distribution of question numbers. They also observed that tooltips display the number of exercises when hovered over. Surprisingly, they discovered an excessive number of exercises in the second exercise and carefully reviewed the content. E1 emphasized, "Teachers should include an appropriate number of questions. Though these questions are most simple, an excessive quantity poses serious problems. It creates high learning pressure and imposes a heavy cognitive load, which hinders effective teaching."

By observing and clicking on the distribution shown in the second set of statistical charts within each exercise, E2 commented, "When designing exercises for first-year students in a C language programming course, it is crucial to maintain a balanced distribution of question types and quantity. This includes multiple-choice, fill-in-the-blank, true or false, and programming tasks. The chart and interaction effectively illustrate the combination of question types based on each chapter's characteristics. For example, in the initial chapters that focus on small code snippets and specific language features, the first three exercises primarily consist of true or false and fill-in-the-blank questions targeting specific concepts. Multiple-choice questions can assess a wider range of concepts. Additionally, programming tasks are employed to guide students through a step-by-step learning process. Notably, when observing the chart across exercises, I also noticed a gradual shift towards programming tasks becoming the dominant question type, which aligns with the nature of a C language programming course."

Distribution of Exercises with Explicit Relationships After assessing the concept distribution, experts further explored the explicit relationships among exercises using the correlation depicted in Fig. 1E(T3). This examination led to the identification of three types of associations: pre-question, similar question, and post-question. For example, as shown in Fig. 4, the central question of this case study involves replacing all English alphabets in the plaintext with the next letter in the alphabet, while converting lowercase letters to uppercase and uppercase letters to lowercase. This task primarily involves branch control and loop control, as demonstrated in the overview and correlation view. Additionally, it encompasses the concept of ASCII conversion. From the correlation view, E3 observed, "Some questions focus on ASCII as the main concept, and they should be introduced prior to this question as prerequisite knowledge. Furthermore, the color-coded

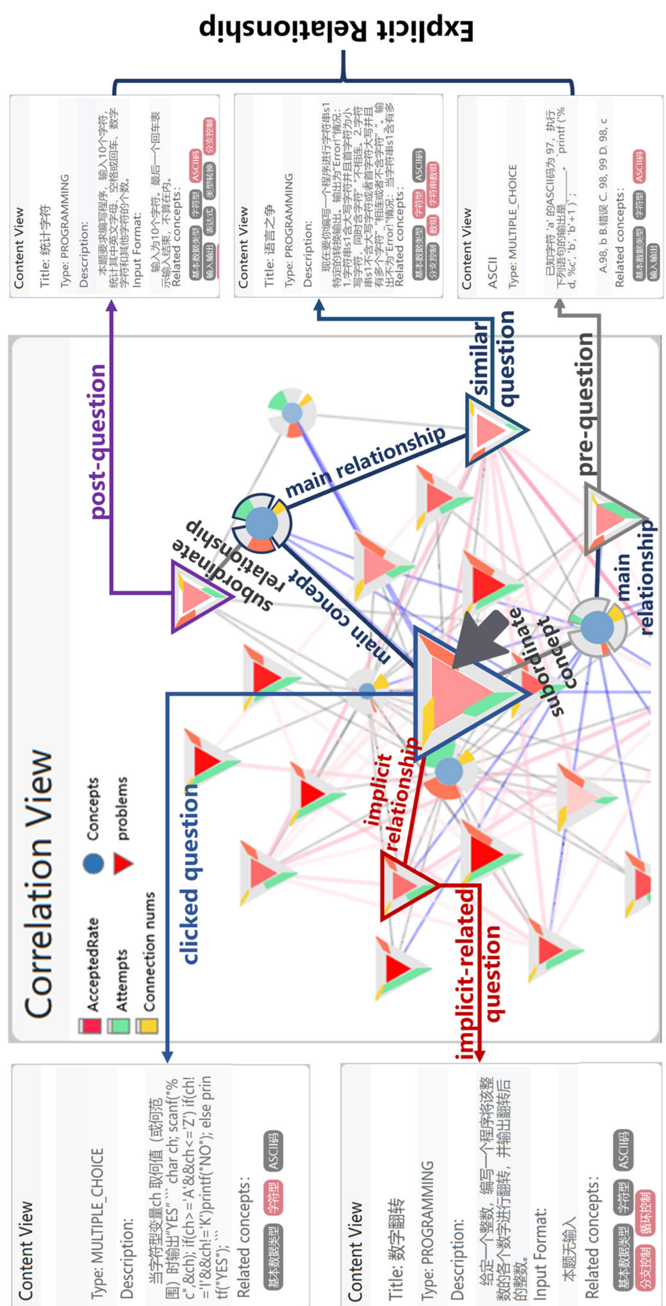

Fig. 4 Explicit relationships derived from concepts among exercises and the implicit relationships inferred from students' performance

connections facilitate the identification of another question that also emphasizes branch control and loop control, indicating that these questions can be grouped together. Moreover, certain questions consider branch control as a subordinate relationship, such as exercises that involve counting the number of characteristics. These exercises are more complex and can be addressed after mastering the fundamental exercises."

6.2 Multi-scale behavior evaluation

Evaluation of Students Behavior. Experts first examined the multidimensional performance of students and analyzed the feedback obtained (T2). They utilized interactive features such as dragging, zooming, and hovering over elements to explore the visualization. During this process, E3 easily identified students requiring attention. She pointed out, "The color contrast was distinct for me to notice that there are some issues with student 1. While hovering over the glyph representing student 1, I can access comprehensive information via the toolkit, including their performance on each exercise/question, which was really helpful. Notably, the parallel coordinate plot clearly depicted fluctuations in accuracy, pinpointing notable periods, such as the fourth to sixth exercises and the ninth to eleventh exercises, during which student 1 achieved zero accuracies. These observations suggested potential knowledge gaps or difficulties in understanding the covered material." Furthermore, E3 remarked, "The student's accuracy gradually changed over time, implying a gradual improvement or decline in the student's understanding and performance. This may also imply a learning curve, with initial struggles followed by a better grasp of concepts, or it could indicate inconsistent progress or engagement throughout the course. Moreover, despite having the fewest attempts, student 1 achieved a passing rate of nearly 0.6. This suggests a certain level of comprehension, although the limited number of attempts could reflect a lack of engagement or participation." In addition to students with a low best score rate, E2 noticed that Student 2 and Student 20 had lower accuracy compared to others, despite achieving higher best score rates. She commented, "These students might have faced challenges in understanding the concepts at first, but their performance improved steadily as they made repeated attempts. Typically, these students are known for their diligence and hard work, and they would benefit from receiving personalized guidance and support. "

After analyzing individual behavior, experts utilized lass circles in Fig. 1D to select user-defined groups for evaluation and comparison. E3 mentioned, "By grouping students, I can analyze patterns and trends, which helps me provide targeted support and customized resources for each group's specific needs. For example, when I looked at individual performance, I quickly noticed issues with students from the northeast. When I selected them, I saw changes in their question behavior and a line appearing in the parallel coordinate plots, showing their similarities. Then I selected more groups and found that the group with students from the center and the previous group were clearly different. This group had fewer attempts and higher score rates, suggesting that they may have a deeper understanding of the material and a higher level of mastery. Based on this, we can encourage collaboration and peer learning, fostering interaction, knowledge sharing, and support within and across different groups, thereby cultivating a collaborative learning environment." Besides, E2 highlighted the system's capability to facilitate behavior analysis and comparison across different classes, while E1 noted its effectiveness in assessing teachers' performance(T2).

Evaluation of exercise and concept Behavior. After analyzing the student behavior, experts delved deeper into the parallel coordinate plot and analyzed the behavior of each exercise and concept. E2 observed that the parallel coordinate plot corresponded with the click action, displaying exercise behavior initially and transitioning to question behavior upon clicking. She noted. "The difference between question types is pretty obvious. Most people only attempt little times in true or false, fill-in-the-blank questions, and multiple-choice questions. However, the situation of programming tasks is pretty complex. Besides, the color and size difference of exercises, questions, and concepts also indicate their appearance. All of them can help me quickly pinpoint specific exercises or concepts that require additional attention or clarification. " Furthermore, by exploring the behavior of each chapter depicted in Fig. 1C_1, experts quickly gained an overview of the course's performance and changes. E1 remarked, "The system was valuable not only in identifying challenging chapters for students but also in distinguishing between teachers and courses."

Additionally, E4 pointed out that the accuracy and attempts charts in the third and fourth visualizations not only display the changes in minimum, maximum, and average levels but also emphasize the variance. He pointed out, "By examining the minimum and maximum scores, I can identify the range of achievement levels and understand the extremes of student performance. Besides, the average score can serve as a measure of central tendency, allowing for an understanding of the overall level of attainment and enabling comparisons against benchmarks or standards. The variance, on the other hand, indicates the degree of dispersion or variability in student grades, highlighting the spread of scores and the diversity of performance within the student group. Moreover, the vertical alignment of exercises and questions enables an overview of each question's performance and facilitates comparisons across questions."

Furthermore, E1 pointed out that in certain exercises, they were able to identify underlying relationships with other exercises through the correlation view. This can be seen in the exercises depicted in Fig. 4, where a colored association indicates these relationships. Further exploration revealed that these exercises are connected through the ASCII concept. E2 added, "By examining the parallel coordinate plots, I observed that most individuals struggle with both the former and latter questions. Notably, while the latter question can be approached using various solutions, it is also possible to solve it using ASCII. Although these two questions may not have an obvious relationship, they have the potential for further exploration and investigation. The implicit relationships do reveal some interesting facts(T3)."

6.3 Expert interview

Based on the previous exploration, we conducted in-depth interviews with domain experts to gather their insights and invited 6 teachers with relevant backgrounds to participate in a questionnaire, aiming to validate the effectiveness and practicality of our system.

The domain experts confirmed that our system was well-designed, considering the characteristics of exercises, and praised its user-friendly interface. E1 mentioned, "The system visualizes content and student behavior at multiple levels. Through the visual design, I can observe overall behavior and identify notable events intuitively. I would like to apply this system to other subjects to evaluate the behavior of different courses or teachers, which may reveal interesting phenomena." E2 highlighted the effectiveness of our design in showcasing differences and associations, making it easier to analyze exercise appropriateness and identify potential issues. She emphasized, "Specifically, the system offers valuable suggestions for exercise types, question types, and quantities. Furthermore, it allows me to adjust exercise distributions and prompts teachers to focus more on concepts that require attention. However, the current data processing is semi-automatic, which can be improved for greater convenience. Automation should be further developed in the system." As a teaching expert, E3 highlighted that the system's interactions, such as selection, tooltips, and highlighting, met her needs for quickly identifying notable students and accessing relevant information to provide further assistance. She also suggested, "The system serves as a convenient tool for monitoring student behavior based on exercise performance. If it could integrate other behaviors, such as forum posts or video-watching, and analyze a larger sample size, I could explore more possible reasons and consider appropriate measures." In contrast to the previous experts, E4, the analyst, noticed the variety of statistical data within the system, which was well-organized. He commented, "During online learning exercises, the number of attempts provides valuable information about students' learning progress and perseverance. However, it may not provide a complete picture of a student's learning journey or their comprehension of the material. Therefore, it is crucial to consider other factors."

Overall, the expert feedback highlighted the system's strengths, including its visual design, effectiveness in analyzing exercise appropriateness, provision of valuable suggestions, and support for monitoring and assisting students. Furthermore, the experts expressed interest in expanding the system's capabilities to encompass additional behaviors and subjects, thus enabling a more comprehensive evaluation of student behavior and informing appropriate interventions.

After interviews, we surveyed 10 teachers from diverse backgrounds to validate the effectiveness and practicality of ExeVis. Participants had the normal or corrected-to-normal vision and received an overview of the study objectives beforehand. The questionnaire, designed following recommended guidelines (Lund 2001; Rossi et al. 2018), used a 7-point Likert scale, ranging from "totally disagree" (1) to "totally agree" (7). The survey consisted of three steps and lasted approximately 30 min. Participants were introduced to ExeVis and received a 10-minute training session. They then had 15 min to explore the system freely before completing the questionnaire in the final 5 min.

The results from Fig. 5 indicate that participants had a generally positive perception of the system, with average ratings ranging from 4.99 to 5.77 across all seven questions. Specifically, Q1–Q3 received an average rating of 5.65, indicating consistent positive evaluations and overall satisfaction with the system. In terms of the visual components (Q4–Q7), the system received an average rating of 5.20 out of 7, suggesting that participants found the visual design engaging and effective in conveying information. However, participants expressed some challenges in identifying relationships between exercises (Q6) due to their

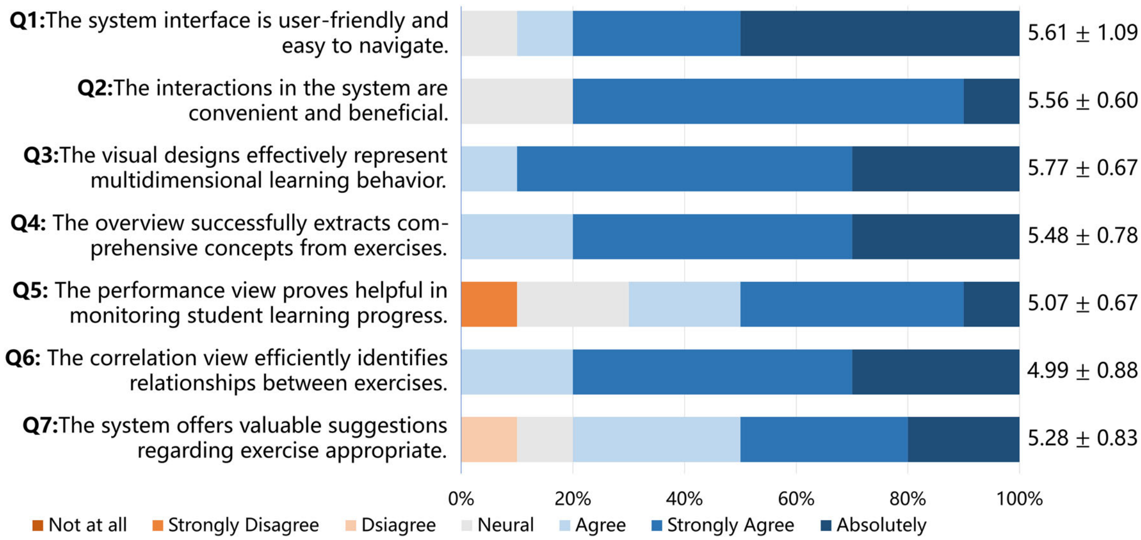

Fig. 5 The description and behavior of questionnaire, while the rightmost column represents Mean ± SD

complexity and abstract nature. Overall, the questionnaire responses reflect a positive performance of the system.

6.4 Discussion

Based on the case study and expert interview, we have identified limitations and corresponding reflections on the system:

The current implementation relies on a combination of automatic and semi-automatic methods, leading to potential inconvenience. Moreover, the user-defined group selection approach may introduce bias or subjectivity. To overcome these limitations, further research is needed to enhance automation capabilities and develop a more robust and data-driven approach to group selection. This entails the development of advanced algorithms and techniques that automate data processing, visualization generation, and interaction mechanisms. A fully automated system would provide users with a seamless and streamlined experience, enabling them to focus on insights and analysis.

Secondly, to overcome this limitation, we suggest expanding the analysis to include more variables and dimensions. This can be achieved by integrating additional data sources, such as online forums and video-watching behavior, to gain a more comprehensive understanding of student engagement and learning patterns.

Additionally, the use of a single dataset in our study limits our ability to capture variations across different educational contexts. To obtain a more comprehensive understanding of the visualization system's effectiveness and applicability, we must extend our analysis to multiple datasets encompassing diverse educational contexts.

By addressing these limitations and incorporating automated techniques, comprehensive analysis, and multiple datasets, we can advance the capabilities and insights provided by ExeVis. This will contribute to a more effective and user-friendly tool for supporting educational decision-making and improving online learning outcomes.

7 Conclusion

In this paper, we present ExeVis, a comprehensive visualization system designed for analyzing exercises in online learning. It introduces several innovative visual encodings and multiple interactive visualization views to help educators gain valuable insights from the distribution of exercises and learner behavior. Furthermore, a rich set of interactions are designed to focus on local areas of interest, and a set of comparison metrics are measured and visually presented. Case studies based on real-world datasets and interviews with domain experts are provided to further demonstrate the effectiveness of our system in evaluating the appropriateness of exercises across different dimensions and enabling educators to adjust their teaching methods accordingly.

Acknowledgements This work was supported in part by the National Natural Science Foundation of China (62277013, 62177040,U22A2032), National Key R &D Program of China (Grant No. 2022YFE0137800), National Statistical Science Research Project (2022LY099), Zhejiang Provincial Science and Technology Program in China (No. 2021C03137), Zhejiang Provincial Science and Technology Plan Project (2023C01120) Public Welfare Plan Research, and Project of Zhejiang Provincial Science and Technology Department (LTGG23H260003).

References

Abdi H, Williams LJ (2010) Principal component analysis. Wiley Interdiscip Rev Comput Stat 2(4):433–459

Ahadi A, Lister R, Vihavainen A (2016) On the number of attempts students made on some online programming exercises during semester and their subsequent performance on final exam questions. In: Proceedings of the 2016 ACM conference on innovation and technology in computer science education, pp 218–223

Anderson JR (1996) The architecture of cognition

Arruarte J, Larrañaga M, Arruarte A, Elorriaga JA (2021) Measuring the quality of test-based exercises based on the performance of students. Int J Artif Intell Educ 31:585–602

Barba PG, Malekian D, Oliveira EA, Bailey J, Ryan T, Kennedy G (2020) The importance and meaning of session behaviour in a MOOC. Comput Educ 146:103772

Becht E, McInnes L, Healy J, Dutertre C-A, Kwok IW, Ng LG, Ginhoux F, Newell EW (2019) Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol 37(1):38–44

Bransford E, John D, Brown E, Ann L, Cocking E, Rodney R (2000) How people learn: brain, mind, experience, and school, expanded. National Academy Press

Brown PC, Roediger HL III, McDaniel MA (2014) Make it stick: the science of successful learning. Harvard University Press

Chen Y-N, Huang Y, Kong S-Y, Lee L-S (2010) Automatic key term extraction from spoken course lectures using branching entropy and prosodic/semantic features. In: 2010 IEEE spoken language technology workshop. IEEE, pp 265–270

Chen Q, Chen Y, Liu D, Shi C, Wu Y, Qu H (2015) PeakVizor: visual analytics of peaks in video clickstreams from massive open online courses. IEEE Trans Visual Comput Graph 22(10):2315–2330

Chen Q, Yue X, Plantaz X, Chen Y, Shi C, Pong T-C, Qu H (2018) ViSeq: visual analytics of learning sequence in massive open online courses. IEEE Trans Vis Comput Graph 26(3):1622–1636

Chen Y, Chen Q, Zhao M, Boyer S, Veeramachaneni K, Qu H (2016) DropoutSeer: visualizing learning patterns in massive open online courses for dropout reasoning and prediction. In: 2016 IEEE conference on visual analytics science and technology (VAST). IEEE, pp 111–120

Chu H-C, Hwang G-J, Tseng JC, Hwang G-H (2006) A computerized approach to diagnosing student learning problems in health education. Asian J Health Inf Sci 1(1):43–60

Cooper M, Zhao J, Bhatt C, Shamma DA (2018) Moocex: exploring educational video via recommendation. In: Proceedings of the 2018 ACM on international conference on multimedia retrieval, pp 521–524

Dai Y, Asano Y, Yoshikawa M (2016) Course content analysis: an initiative step toward learning object recommendation systems for MOOC learners. International Educational Data Mining Society

Davis MC, Duryee LA, Schilling AH, Loar EA, Hammond HG (2020) Examining the impact of multiple practice quiz attempts on student exam performance. J Educ Online 17(2):2

Devlin J, Chang M-W, Lee K, Toutanova K (2018) Bert: pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805

Dunlosky J, Rawson KA, Marsh EJ, Nathan MJ, Willingham DT (2013) Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. Psychol Sci Public Interest 14(1):4–58

Emmons SR, Light RP, Börner K (2017) MOOC visual analytics: empowering students, teachers, researchers, and platform developers of massively open online courses. J Am Soc Inf Sci 68(10):2350–2363

Ferguson R (2012) Learning analytics: drivers, developments and challenges. Int J Technol Enhanc Learn 4(5–6):304–317

Ferguson R, Clow D (2015) Examining engagement: analysing learner subpopulations in massive open online courses (MOOCS). In: Proceedings of the fifth international conference on learning analytics and knowledge, pp 51–58

Firat EE, Denisova A, Wilson ML, Laramee RS (2022) P-Lite: a study of parallel coordinate plot literacy. Vis Inform 6(3):81–99. https://doi.org/10.1016/j.visinf.2022.05.002

Gove R, Cadalzo L, Leiby N, Singer JM, Zaitzeff A (2022) New guidance for using t-SNE: alternative defaults, hyperparameter selection automation, and comparative evaluation. Vis Inform 6(2):87–97. https://doi.org/10.1016/j.visinf.2022.04.003

Grover S, Pea R, Cooper S (2014) Promoting active learning & leveraging dashboards for curriculum assessment in an OpenEdX introductory CS course for middle school. In: Proceedings of the first ACM conference on learning@ scale conference, pp 205–206

Han D, Pan J, Pan R, Zhou D, Cao N, He J, Xu M, Chen W (2022) iNet: visual analysis of irregular transition in multivariate dynamic networks. Front Comput Sci 16(2):162701. https://doi.org/10.1007/s11704-020-0013-1

Han R, Peng T, Yang C, Wang B, Liu L, Wan X (2023) Is information extraction solved by ChatGPT? An analysis of performance, evaluation criteria, robustness and errors

Han F, Veeramachaneni K, O'Reilly U-M (2013) Analyzing millions of submissions to help mooc instructors understand problem solving. In: NIPS workshop on data driven education. Curran Associates, pp 1–5
Hattie J, Timperley H (2007) The power of feedback. Rev Educ Res 77(1):81–112
Helic D, Maurer H, Scherbakov N (2000) Web based training: what do we expect from the system. In: Proceedings of ICCE, pp 1689–1694
Huang N-F, Hsu H-H, Chen S-C, Lee C-A, Huang Y-W, Ou P-W, Tzeng J-W (2017) Videomark: a video-based learning analytic technique for MOOCS. In: 2017 IEEE 2nd international conference on big data analysis (ICBDA). IEEE, pp 753–757
Huang J, Huang A, Lu O, Tseng H-C, Yang S (2016) Learning dashboard: visualization of learning behaviour in MOOCS. In: The international workshop on technology-enhanced collaborative learning (TECL 2016) in conjunction with CRIWG/ CollabTech 2016, vol 1, p 25
Huang L, Xu M, Chen Z, Liu F (2019) Syllabus design for teacher education MOOCS (massive open online courses): a mixed methods approach. In: Technology in education: pedagogical innovations: 4th international conference, ICTE 2019, Guangzhou, China, March 15–17, 2019, revised selected papers 4, pp 156–168. Springer
Koedinger KR, Kim J, Jia JZ, McLaughlin EA, Bier NL (2015) Learning is not a spectator sport: doing is better than watching for learning from a MOOC. In: Proceedings of the second ACM conference on learning@ scale, pp 111–120
Kui X, Liu N, Liu Q, Liu J, Zeng X, Zhang C (2022) A survey of visual analytics techniques for online education. Vis Inform 6(4):67–77. https://doi.org/10.1016/j.visinf.2022.07.004
Li B, Fang G, Yang Y, Wang Q, Ye W, Zhao W, Zhang S (2023a) Evaluating ChatGPT's information extraction capabilities: an assessment of performance. Calibration, and faithfulness, explainability
Li J, Li H, Pan Z, Pan G (2023b) Prompt ChatGPT in MNER: improved multimodal named entity recognition method based on auxiliary refining knowledge from ChatGPT
Liu C, Kim J, Wang H-C (2018) Conceptscape: collaborative concept mapping for video learning. In: Proceedings of the 2018 CHI conference on human factors in computing systems, pp 1–12
Loper E, Bird S (2002) Nltk: the natural language toolkit. arXiv preprint arXiv:0205028
Lund AM (2001) Measuring usability with the use questionnaire. Usability Interface 8(2):3–6
Maaten L, Hinton G (2008) Visualizing data using t-SNE. J Mach Learn Res 9(11):2579–2605
Mayer R (2005) The Cambridge handbook of multimedia learning: contributors
Mazza R, Dimitrova V (2007) CourseVis: a graphical student monitoring tool for supporting instructors in web-based distance courses. Int J Hum Comput Stud 65(2):125–139
Means B, Toyama Y, Murphy R, Bakia M, Jones K (2009) Evaluation of evidence-based practices in online learning: a meta-analysis and review of online learning studies
Merceron A, Yacef K (2005) Educational data mining: a case study. In: AIED, pp 467–474
Mubarak AA, Ahmed SA, Cao H (2021) MOOC-ASV: analytical statistical visual model of learners' interaction in videos of MOOC courses. Interact Learn Environ 25:1–16
Munzner T (2009) A nested model for visualization design and validation. IEEE Trans Vis Comput Graph 15(6):921–928
Nesbit JC, Adesope OO (2006) Learning with concept and knowledge maps: a meta-analysis. Rev Educ Res 76(3):413–448
OpenAI (2023) Openai api. https://platform.openai.com/
Pan L, Li C, Li J, Tang J (2017) Prerequisite relation learning for concepts in MOOCS. In: Proceedings of the 55th annual meeting of the association for computational linguistics (volume 1: long papers), pp 1447–1456
Pellegrino JW (2002) Knowing what students know. Issues Sci Technol 19(2):48–52
Peng Y, Fan X, Chen R, Yu Z, Liu S, Chen Y, Zhao Y, Zhou F (2023) Visual abstraction of dynamic network via improved multi-class blue noise sampling. Front Comput Sci 17(1):165–179
Romero C, Ventura S (2010) Educational data mining: a review of the state of the art. IEEE Trans Syst Man Cybern C (Appl Rev) 40(6):601–618
Rossi PH, Lipsey MW, Henry GT (2018) Evaluation: a systematic approach. Sage Publications
Schwab M, Strobelt H, Tompkin J, Fredericks C, Huff C, Higgins D, Strezhnev A, Komisarchik M, King G, Pfister H (2016) booc.io: an education system with hierarchical concept maps and dynamic non-linear learning plans. IEEE Trans Vis Comput Graph 23(1):571–580
Stephens-Martinez K, Hearst MA, Fox A (2014) Monitoring MOOCS: which information sources do instructors value? In: Proceedings of the First ACM conference on learning@ scale conference, pp 79–88
Tomkins S, Ramesh A, Getoor L (2016) Predicting post-test performance from online student behavior: a high school MOOC case study. International Educational Data Mining Society
Tukey JW et al (1977) Exploratory data analysis, vol 2. Reading, MA
Vranic M, Pintar D, Skocir Z (2007) The use of data mining in education environment. In: 2007 9th international conference on telecommunications. IEEE, pp 243–250
Wang X, Wu Z, Huang W, Wei Y, Huang Z, Xu M, Chen W (2023) VIS+ AI: integrating visualization with artificial intelligence for efficient data analysis. Front Comput Sci. https://doi.org/10.1007/s11704-023-2691-y
Wei H-C, Chou C (2020) Online learning performance and satisfaction: do perceptions and readiness matter? Distance Educ 41(1):48–69
Wei X, Cui X, Cheng N, Wang X, Zhang X, Huang S, Xie P, Xu J, Chen Y, Zhang M et al (2023) Zero-shot information extraction via chatting with ChatGPT. arXiv preprint arXiv:2302.10205
Wiliam D (2011) What is assessment for learning? Stud Educ Eval 37(1):3–14
Yang W, Liu M, Wang Z, Liu S (2024) Foundation models meet visualizations: Challenges and opportunities. Computational Visual Media. arxiv.org/abs/2310.05771
Zhang S, Meng X, Liu C, Zhao S, Sehgal V, Fjeld M (2019) Scaffomapping: assisting concept mapping for video learners. In: Human-computer interaction—INTERACT 2019: 17th IFIP TC 13 international conference, Paphos, Cyprus, September 2–6, 2019, Proceedings, Part II 17. Springer, pp 314–328

Zhang G, Zhu Z, Zhu S, Liang R, Sun G (2022) Towards a better understanding of the role of visualization in online learning: a review. Vis Inform

Zhao B, Lin S, Luo X, Xu S, Wang R (2017) A novel system for visual navigation of educational videos using multimodal cues. In: Proceedings of the 25th ACM international conference on multimedia, pp 1680–1688

Zheng S, Rosson MB, Shih PC, Carroll JM (2015) Understanding student motivation, behaviors and perceptions in MOOCS. In: Proceedings of the 18th ACM conference on computer supported cooperative work & social computing, pp 1882–1895

Zhong Q, Tang Y (2020) An attention-based BiLSTM-CRF for Chinese named entity recognition. In: 2020 IEEE 5th international conference on cloud computing and big data analytics (ICCCBDA). IEEE, pp 550–555

Zhou Z, Shi C, Shen X, Cai L, Wang H, Liu Y, Ying Z, Chen W (2020) Context-aware sampling of large networks via graph representation learning. IEEE Trans Vis Comput Graph PP:1. https://doi.org/10.1109/TVCG.2020.3030440

Zhou Z, Sun L, Yu W, Liu Y, Xiang Z, Wang Y, Chen W (2022) IMGC: interactive multiple graph clustering with constrained Laplacian rank. IEEE Trans Hum Mach Syst PP:1–11. https://doi.org/10.1109/THMS.2022.3227181

Zimet GD, Dahlem NW, Zimet SG, Farley GK (1988) The multidimensional scale of perceived social support. J Pers Assess 52(1):30–41