SBGTool v2.0: An Empirical Study

Contributor: Zeynab (Artemis) Mohseni, Rafael M. Martins, Italo Masiello

SBGTool v2.0 differs from SBGTool due to design changes made in response to teacher suggestions, the addition of sorting options to the dashboard table, the addition of a dropdown component to group the students into classrooms, and improvement in some visualizations. By applying SBGTool v2.0, teachers may compare the outcomes of individual students inside a classroom, determine which subjects are the most and least difficult over the period of a week or an academic year, identify the numbers of correct and incorrect responses for the most difficult and easiest subjects, categorize students into various groups based on their learning outcomes, discover the week with the most interactions for examining students' engagement, and find the relationship between students' activity and study success.

1. Introduction

In a classroom, students interact and discuss directly with teachers and with each other, allowing them to gain an understanding of their performance. However, in a digital environment, students may struggle to determine if their work is on par with that of their peers or whether it meets the teacher's expectations. Establishing groups that allow students to engage in constructive conversations, reflect on their approaches to the subject of study, and move into deeper learning is an issue that teachers face in physical and digital classrooms [1]. Since each student may participate in many activities over long periods of time, datasets that record such activities are usually high-dimensional, large, and complex.

Learning analytics (LA) has evolved to help in the interpretation of educational data [2]. Siemens et al. [3] (p. 1) define LA as "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs". To help teachers interpret LA more effectively, data can be visualized in dashboards by visual learning analytics (VLA) tools. Those are defined by Schwendimann et al. [4] (p. 8) as "a single display that aggregates different indicators about learner(s), learning process(es), and/or learning context(s) into one or multiple visualizations". Therefore, VLA can be used to obtain a graphical understanding of what a teacher or a student can do to improve motivation, self-direction, learning effectiveness, student performance, and teacher engagement [5]. In this research, the researchers improved the SBGTool, a similarity-based grouping VLA tool that presented earlier [6], by conducting a user study to learn more about teachers' demands with such a pedagogical tool.

2. Overall Design of SBGTool v2.0

The researchers describe the design of the SBGTool v2.0. The proposed tool is interactive, and the views are coordinated and interconnected, which allows much deeper interactions than simply a set of graphs generated in Excel. Interactive, coordinated views allow for the testing of hypotheses and multilevel exploration of complex data [7]. As can be seen in **Figure 1**, the researchers employed a strategy of increasing details, starting from Key Metrics, followed by Overview and finally Detail. These three levels follow Shneiderman's mantra [8], "overview first, zoom and filter, then details on demand", which drives visual information-seeking behavior and interface design. The most important global information about the dataset is displayed in the *Key Metrics* section (**Figure 1**A), which includes the total numbers of correct and incorrect answers, number of students, questions, and number of answers in the four answer choices (A, B, C, D).

Figure 1. Overview of tabs and components of SBGTool v2.0. (**A**) Key Metrics section, (**B**) Overview section, (**C**) Detail section presented in 2 parts of the tool.

To make it easier to recognize key metrics in different visualizations, the researchers used the following colors: ● dark blue for correct answer, ● red for incorrect answer, ● light blue for option A, ● green for option B, ● pink for option C, and ● orange for option D. Since color blindness affects about 1 in every 20 people, the researchers chose colors that are appropriate for people who are colorblind (**Figure 2**) [9]. The researchers utilized the possibility of choosing the color of the website [9] to examine what the researchers' chosen color palette would seem like to colorblind viewers. The colors in the leftmost column of **Figure 2** are the "true" colors, which are shown in the remaining three columns as they would seem to someone with protanopia, deuteranopia, or tritanopia, i.e., color blindness or less sensitivity to red, green, or blue-yellow light, respectively [9].

The *Overview* section of SBGTool v2.0 (**Figure 1**B) includes a view that displays the total numbers of correct and incorrect answers, the total number of students' answers for the four answer choices, and the total number of correct answers for the four answer choices over time. This view is an overlay of grouped bar and line charts. By selecting a point (week) in the overview, all the key metrics and visualizations in the *Detail* section (**Figure 1**C) of SBGTool v2.0 are updated accordingly. In addition, a range slider on the overview enables the teacher to limit the *x*-axis value within a given range (minimum and maximum values). By moving the mouse in the overview and looking at the number of students' answers and correct answers (in the four answer choices for multiple-choice questions in a given week), the teachers can determine how near the students' learning activities were to a correct response. In order to focus on the students' activity inside a classroom, a class name should be selected from the dropdown component on the top right side of the tool. To reload the entire selected 10,000 samples, the "reset" button on the tool's top right side can be pressed. Likewise, the "learn more" button next to the "reset" button may be used to obtain a general description of the tool that has been developed.

SBGTool v2.0 allows the user to interact with each visualization individually and drill down to more detailed levels of information. Brushing, zooming, and filtering are all supported by the majority of the visualizations. The main purpose of brushing is to emphasize brushed data elements in the various tool views. The user can obtain more information by picking a portion of each graphic and zooming in. The user can also filter the view by clicking on the legend (square or circle) in the right half of each display.

The *Detail* section of the proposed tool (**Figure 1**C) includes a table and two bar charts on the left-hand side, as well as three tabs with specific visualizations to aid in the discovery of insight about various subjects, students' groups with similar activities in terms of the number of correct answers, number of student interactions, number of interactions in different features, and the difference between students' learning activities.

Teachers can obtain thorough information on students' activities and the time they set aside to answer a question by glancing at the table shown in **Figure 1**C. In addition, they can filter and sort the table according to the student ID, date week, date, subject number, user answer, correct answer, result, and answer duration, to have more focused information. The first bar chart shown in **Figure 1**C on the left-hand side depicts the overall percentages of correct and incorrect responses. As can be seen in the bar chart presented in **Figure 1**C, the percentages of correct and incorrect answers for students in class 1 are 70.47% and 29.53%, respectively. The second bar chart on the left-hand side of the tool with multicategory axis type presents the percentages of difficulty and ease in seven subjects. For example, as seen in this bar chart, the percentages of difficulty and ease in subject 1 are 23% and 77%, respectively. By comparing these percentages to those in other subjects, the researchers can determine the most difficult and easiest subjects. In order to determine the percentage of difficulty in different subjects, the researchers use Equation (1) where N_C and N_I are the numbers of correct and incorrect answers, respectively, in each subject. Equation (1) is based on the difficulty index [10] which allows us to calculate the proportion of students who correctly answered the questions belonging to a subject.

$$Difficulty_{Sub.} = (1 - \frac{N_C}{N_C + N_I}) \times 100,$$

(1)

Figure 3 displays the "Students' performance" tab which contains a scatter plot and selections (radio buttons) for rendering a set of features in the scatter plot. By selecting the student ID from the checkbox list, the performance levels are shown (**Figure 3**a). To define the percentage of performance, the researchers apply Equation (2) where N_C and N_I are the numbers of correct and incorrect answers, respectively, for each student.
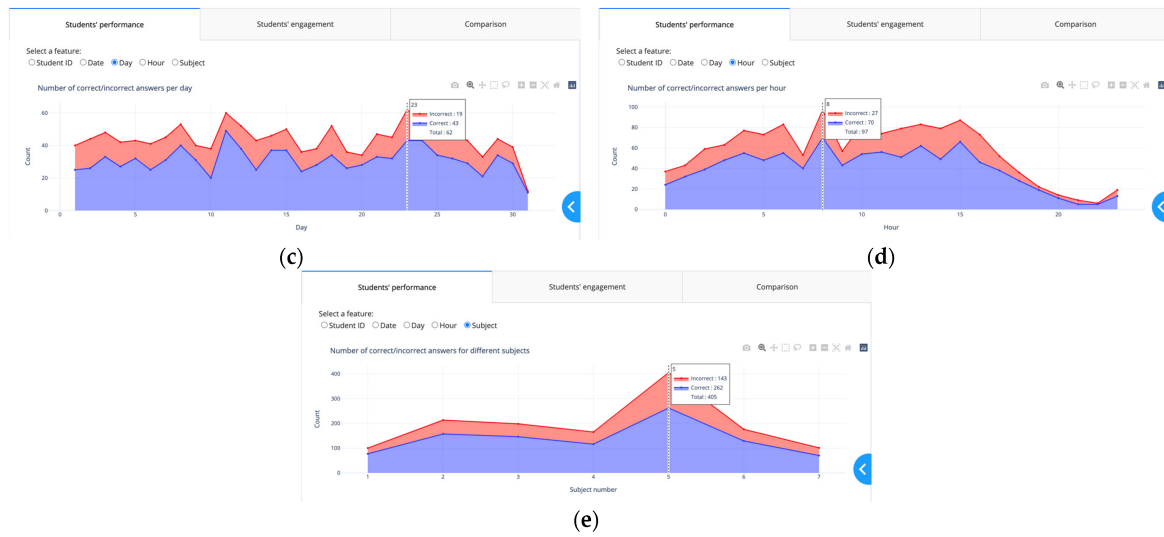
$$Performance = (\frac{N_C}{N_C + N_I}) \times 100,$$ (2)

(**a**)　　　　　(**b**)

(c)

(d)

(e)
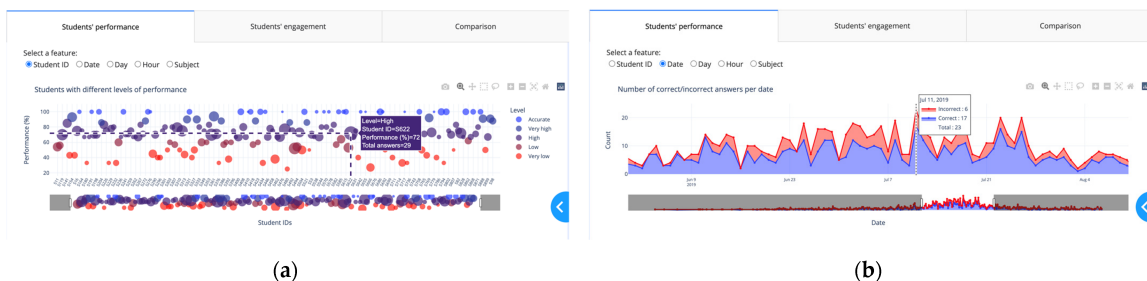
Figure 3. "Students' performance" tab with different checkbox items: (**a**) student ID; (**b**) date; (**c**) day; (**d**) hour; (**e**) subject.

For the color scale, the researchers considered five levels of performance, in increasing order: "Very low", "Low", "High, "Very high", and "Accurate". The researchers used the proportion of performance to group the students into these different levels of performance. For color blindness, the researchers chose the color palette presented in **Figure 4** where the colors in the leftmost column are the "true" colors, which are shown in the remaining three columns as they would seem to someone with protanopia, deuteranopia, or tritanopia. Students with a performance of 100% are in the 'Accurate" level, students with a performance of 86% to 99% are in the "Very high" level, students with a performance of 66% to 85% are in the "High" level, students with a performance of 51% to 65% are in the "Low" level, students with a performance of 1% to 50% are in the "Very low" level. According to the performance levels, the researchers considered "Accurate" and "Very high" levels as high-performing, "High" as average-performing, and "Low" and "Very low" as low-performing.
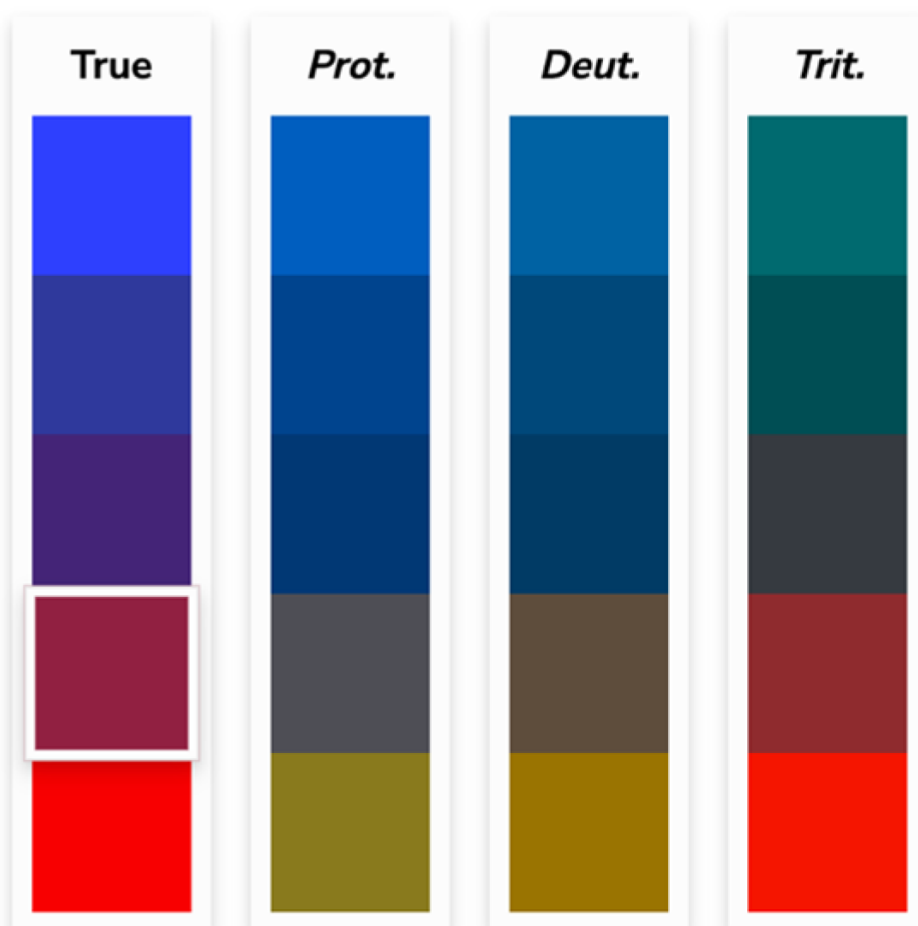
Figure 4. Chosen colors for the five levels of performance in the scatter plot. Reprinted/adapted with permission from Ref. [9]. Copyright 2022, David Nichols.

Teachers can use the visualizations in **Figure 3**b–d to determine the total number of interactions (the total numbers of correct and incorrect answers) and the numbers of correct and incorrect answers for each date, day, and hour.

The numbers of correct and incorrect answers and the total number of interactions for each subject are shown in **Figure 3**e. Since the *x*-axis of the view in the Overview section represents the weekly activity, the visualizations in this tab allow teachers to dig deeper into the numbers of correct and incorrect answers for each feature. The "Students' engagement" tab shown in **Figure 5** displays the students' activity levels. Because the heatmap is arranged from left to right, teachers can quickly identify students who have similar numbers of activities in each of the four answer choices, as well as those who have the most and least interactions with the digital learning material. Students with the most interactions are represented by navy blue, while those with the fewest interactions are represented by light blue. **Figure 6** illustrates the "Comparison" tab that includes two dropdown components and two Sankey diagrams. A Sankey diagram can be thought of as a flow diagram in which the width of arrows is proportional to the flow quantity. In these diagrams, the researchers visualize the contributions to a flow by designating "Student ID" as a source; "Result" as a target; and "Date Week", "Subject number", and "Student answer" as flow volume. Using this tab enables a teacher to compare the students' learning outcomes by choosing their IDs from the dropdown components.
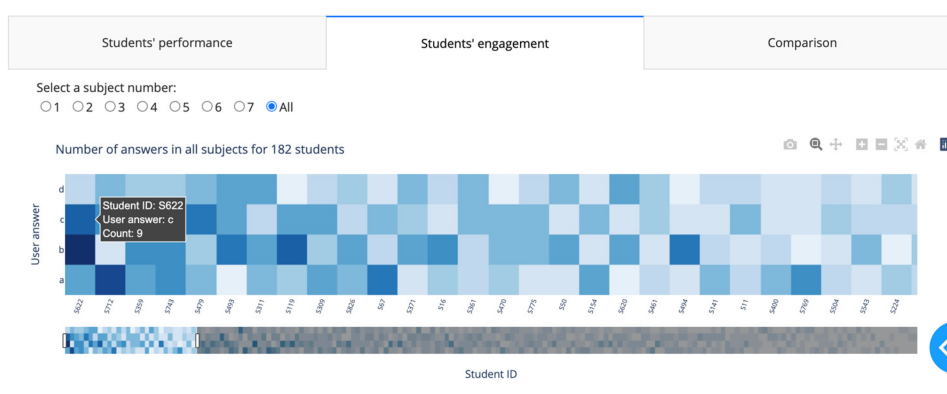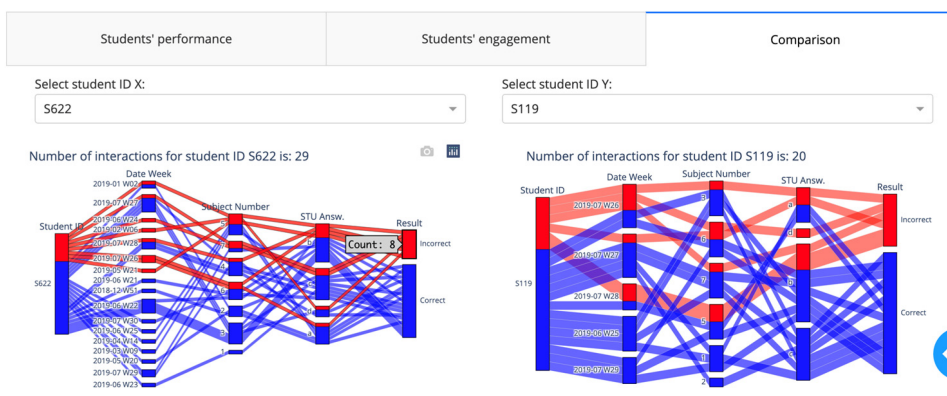

Figure 5. "Students' engagement" tab.

Figure 6. "Comparison" tab.

References

1. Blakelock, J.; Smith, T.E. Distance learning: From multiple snapshots, a composite portrait. Comput. Compos. 2006, 23, 139–161.

2. Aruvee, M.; Ljalikova, A.; Vahter, E.; Prieto, L.; Poom-Valickis, K. Learning analytics to informand guide teachers as designers of educational interventions. In Proceedings of the International Conference on Education and Learning Technologies, Palma, Spain, 2–4 July 2018.

3. Siemens, G.; Baker, R.S.J.d. Learning analytics and educational data mining: Towards communication and collaboration. In Proceedings of the LAK '12: 2nd International Conference on Learning Analytics and Knowledge, Vancouver, BC, Canada, 29 April–2 May 2012; pp. 252–254.

4. Schwendimann, B.A.; Rodríguez-Triana, M.J.; Vozniuk, A.; Prieto, L.P.; Boroujeni, M.S.; Holzer, A.; Gillet, D.; Dillenbourg, P. Perceiving learning at a glance: A systematic literature review of learning dashboard research. IEEE Trans. Learn. Technol. 2016, 10, 30–41.

5. Verbert, K.; Ochoa, X.; Croon, D.; Dourado, R.A.; Laet, T.D. Learning analytics dashboards: The past, the present and the future. In Proceedings of the LAK20: Tenth International Conference on Learning Analytics and Knowledge, Frankfurt, Germany, 23–27 March 2020; pp. 35–40.

6. Mohseni, Z.; Martins, R.M.; Masiello, I. SBGTool: Similarity-Based Grouping Tool for Students' Learning Outcomes. In Proceedings of the 2021 Swedish Workshop on Data Science (SweDS), Växjö, Sweden, 2–3 December 2021

7. Roberts, J.C. State of the Art: Coordinated & Multiple Views in Exploratory Visualization. In Proceedings of the Fifth International Conference on Coordinated and Multiple Views in Exploratory Visualization (CMV 2007), Zurich, Switzerland, 2 July 2007; pp. 61–71.

8. Shneiderman, B. The eyes have it: A task by data type taxonomy for information visualizations. In Proceedings of the 1996 IEEE Symposium on Visual Languages, Boulder, CO, USA, 3–6 September 1996.

9. Nichols, D. Coloring for Colorblindness. Available online: https://davidmathlogic.com/colorblind/#%23648DF9-%23785EF0-%23E21F7F-%23F7833C-%23FFBA21-%2376EA74 (accessed on 10 June 2022).

10. Hingorjo, M.R.; Jaleel, F. Analysis of One-Best MCQs: The Difficulty Index, Discrimination Index and Distractor Efficiency. J. Pak. Med. Assoc. 2012, 62, 142–147.

Retrieved from https://encyclopedia.pub/entry/history/show/62740