

Relationship Between Students' Scores in Weekly Tests and Final Exam

Satomi HAMADA^{a*}, Izumi HORIKOSHI^b & Hiroaki OGATA^b

^aGraduate School of Informatics, Kyoto University, Japan

^bAcademic Center for Computing and Media Studies, Kyoto University, Japan

*hamada.satomi.36y@st.kyoto-u.ac.jp

Abstract: This study investigates the relationship between students' scores in weekly tests and the unit exam. By analyzing score patterns, we found that students who consistently scored high on weekly tests performed better in the unit exam, whereas those who struggled with the unit's contents early tended to score lower. These findings emphasize the importance of using weekly test scores as formative assessments to help students monitor their progress and adjust their learning.

Keywords: Learning Analytics, Weekly Test, Unit Exam, Formative Assessment

1. Introduction

Recent research has highlighted the importance of incorporating summative and formative assessments into classroom assessments (Black & Wiliam, 2018). One of the methods used in formative assessment is a test on the content of the class, and appropriate test design and feedback contribute to students' final grades (Peterson & Siadat, 2009). Many studies of evidence-based formative assessment have been conducted in experimental environments (Morris et al., 2021). However, laboratory and real-world environments are different, which can interfere with the interpretation of the results, and some studies have attempted to verify the pure effect in real-world settings (Greving & Richter, 2018)

In this study, we focused on students' weekly test scores in the real world and examined the relationship between score changes in weekly tests and the unit exam. If this relationship can be clarified, it will be possible to promote students' awareness of their future performance and help them learn faster rather than wait for the unit exam score. Therefore, we pose the following research questions:

RQ1: *What are the patterns of change in students' weekly test scores?*

RQ2: *Do the patterns of change in students' weekly test scores influence their performance on the unit exam?*

RQ3: *Do the patterns of change in students' weekly test scores differ from unit to unit?*

2. Method

The data used in this study comprised scores of weekly tests and unit exams conducted in mathematics classes for 9th grade students in Japan. As shown in Figure 1, the school conducts weekly tests (e.g., May 22, May 30, and June 5) and a one-unit exam for each unit. The weekly test was designed by the teacher in such a way that all students can obtain a perfect score if they understand the teaching and supplementary materials used in the class. Because the perfect scores differ depending on the test, we converted the scores into the correct rate for each weekly test and used them for the analyses. However, because the maximum score for both unit exams was 100 points, the original scores were used for the analysis. Each unit has a set of weekly tests and unit exams, and in this analysis, two units were analyzed: "Figures and Similarities (n=54)" and "Pythagorean Theorem (n=15)."

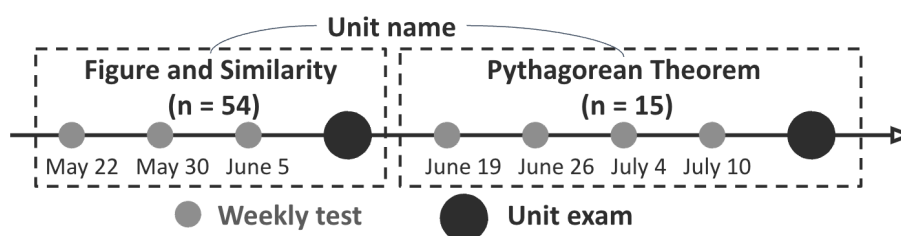


Figure 1. The Flow of Weekly Tests and Unit Exams

3. Results and Discussion

3.1 Analysis 1: The Clustering of Weekly Test Score Changes

Using the datasets of "Figures and Similarities" and "Pythagorean Theorem," we verified whether the patterns of weekly test score changes could be classified. As shown in Figure 1, the number of weekly tests was three times for "Figures and Similarities" and four times for "Pythagorean Theorem." To summarize these two units ($n=69$) and identify the change patterns, DTW K-means, which can be used for different time series, was performed. The results are shown in Figure 2. When the number of clusters was set to three, based on the results of the elbow method, the patterns of weekly test score changes were classified into a high-score keepers that always had a near-perfect score, a falling score group that gradually dropped toward the end of the unit, and a middle score keepers that was always approximately 80%. Among the three clusters revealed in this analysis, no group showed a gradual increase in the scores. This result is consistent with Hirose's (2018) finding that students who fell steeper in their weekly test scores tended to fail in class. In other words, it is essential to improve students' struggles with unit content during the early stages of the unit.

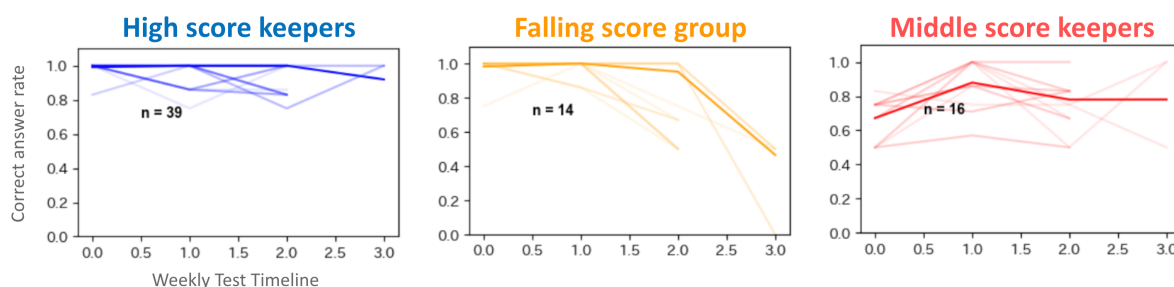


Figure 2. The Results of Clustering Weekly Test Score Changes ($n=69$)

3.2 Analysis 2: Trends in Unit Exam Scores by Cluster

To analyze whether the trend of unit exam scores differed for each cluster of weekly test scores classified in Section 3.1, the unit exam scores of students belonging to that cluster were plotted as box plots. The results are shown in Figure 3. In both units, it was confirmed that the high score keepers, who was always close to the perfect score, tended to obtain the highest score in the unit test, followed by the falling score group and the middle score keepers. Furthermore, an ANOVA was performed on the unit exam scores for each cluster, which revealed significant differences in only "Figure and Similarities" ($p=0.001$, $p<0.05$) with an F value of 7.333. The results of the Tukey test for post-hoc comparison showed a statistically significant difference between the high- and middle-score keepers. In addition, it may not confirm significant differences in the "Pythagorean Theorem" due to the sample size; therefore, future studies must be conducted with larger sample sizes. Similar to the results mentioned in Section 3.1, students' instability in the early stages of the unit affect the final test result. This analysis helps students estimate their final grades based on the weekly test before the final test results are returned. In other words, information on weekly test score changes could help students become aware of improving their learning as a formative assessment.

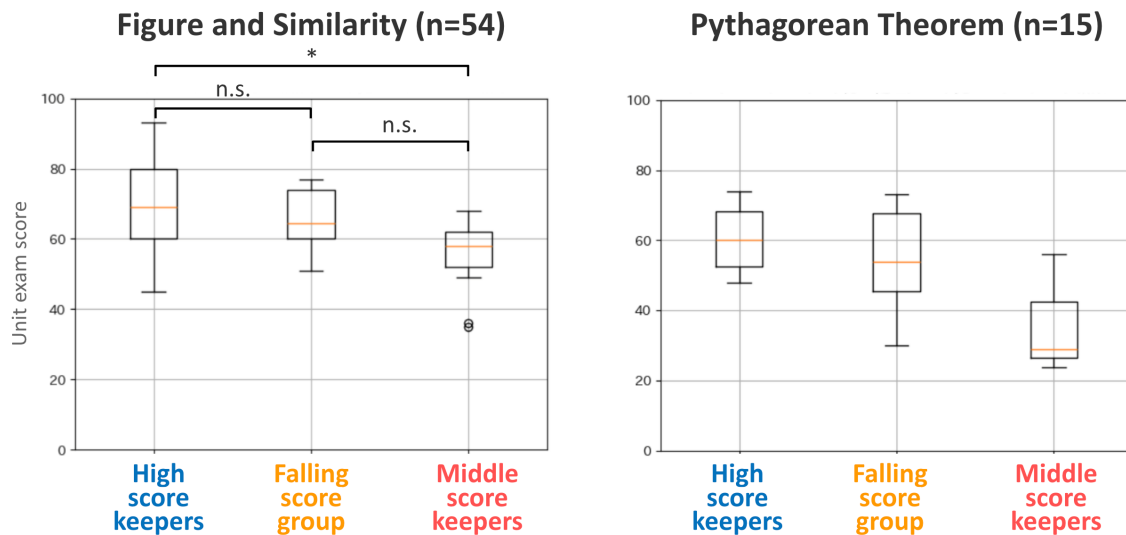


Figure 3. The Boxplot of Unit Exam Scores by Cluster

3.3 Analysis 3: Changes in the Cluster in both Units

To analyze whether the cluster to which students belong can change depending on the unit, we used the data of students who recorded the weekly test scores and unit exam scores of both units (n=12) and plotted the changes in their clusters on a Sankey-Diagram. The results are shown in Figure 4. Three students remained in the high score keepers, one in the falling-score group, and two in the middle score keepers, with half of the students maintaining a similar trend in different units. The other half of the students changed from high score keepers to the falling score group and from middle score keepers to high score keepers. These changes may be due to the students' strengths and weaknesses in each unit or may result from the students' improvement actions. In this study, we only identified the changes in the cluster, and in the future, it is necessary to analyze the process of why and what kind of change occurred.

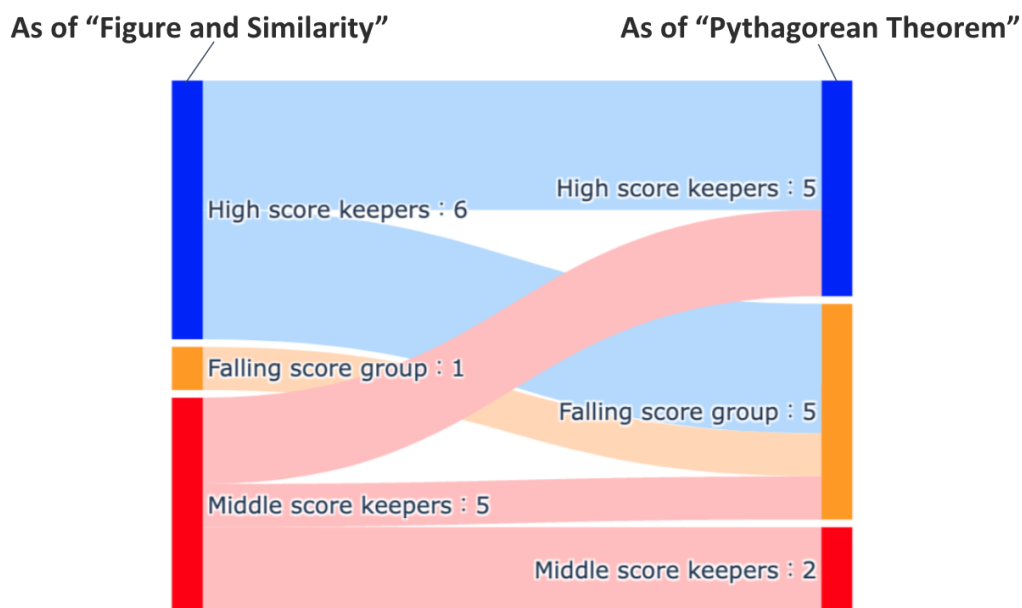


Figure 4. The Sankey-Diagram of Changes in Clusters for both Units (n=12)

4. Conclusion and Future Work

This study analyzed students' weekly test scores in the real world and examined the relationship between changes in weekly tests and unit exam scores. The results showed that the patterns of weekly test score changes were classified into three clusters: high score keepers, falling score group, and middle score keepers. The unit exam scores were high in this order. The results show that it is essential to eliminate students' instability at an early stage in the unit rather than just before the final test. In addition, changes in weekly test scores can predict the final grade, and this information can be used to encourage students to learn better at an early stage, which can contribute to their formative assessment.

However, as a limitation of the study, the data used in this analysis are limited, and it is necessary to clarify this trend by using more data in the future. In addition, it is necessary to expand the analysis of the changes in the cluster of each unit, as mentioned in Section 3.3, using trace data to analyze the factors that led to the change. Thus, we hope to analyze the provision of advice with specific behavioral indicators to improve student learning behavior.

Acknowledgments

This study was supported by the CSTI SIP program (grant number: JPJ012347).

References

- Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, 25(6), 551–575. <https://doi.org/10.1080/0969594X.2018.1441807>
- Greving, S., & Richter, T. (2018). Examining the testing effect in university teaching: Retrievability and question format matter. *Frontiers in Psychology*, 9, 2412.
- Hirose, H. (2018). Difference between successful and failed students learned from analytics of weekly learning check testing. *Information Engineering Express*, 4(1), 11–21.
- Morris, R., Perry, T., & Wardle, L. (2021). Formative assessment and feedback for learning in higher education: A systematic review. *Review of Education*, 9(3). <https://doi.org/10.1002/rev3.3292>
- Peterson, E., & Siadat, M. V. (2009). Combination of formative and summative assessment instruments in elementary algebra classes: A prescription for success. *Journal of Applied Research in the Community College*, 16, 92–102.