2023-10-10

Interconnectedness Of Geometric, Linguistic, And Algebraic Thinking In Student Performance Measures: An Association Rules Approach

Bence SIPOS
Budapest University of Technology and Economics, Hungary;MTA–ELTE Theory of Learning Mathematics Research Group, bence.sipos.sb@gmail.com

Brigitta SZILÁGYI
Budapest University of Technology and Economics, Hungary;Corvinus University of Budapest;MTA–ELTE Theory of Learning Mathematics Research Group, szilagyibr@gmail.com

INTERCONNECTEDNESS OF GEOMETRIC, LINGUISTIC AND ALGEBRAIC THINKING IN STUDENT PERFORMANCE MEASURES: AN ASSOCIATION RULES APPROACH

Bence Sipos[1]
Budapest University of Technology and Economics
Budapest, Hungary
MTA–ELTE Theory of Learning Mathematics Research Group

Brigitta Szilágyi
Budapest University of Technology and Economics
Budapest, Hungary
MTA–ELTE Theory of Learning Mathematics Research Group

Conference Key Areas: *Mentoring and Tutoring, Fundamentals of Engineering: Mathematics and the Sciences*
Keywords: *Multidisciplinary, Association-rules, assessments, Interconnectedness*

ABSTRACT

Assessing student performance is crucial in education for evaluating knowledge acquisition and competency development. Traditional grading systems often overlook the interconnectedness of learning domains, which can provide valuable insights into student understanding. This study investigates the associations between geometric, linguistic, and algebraic thinking and their impact on student performance measures and grading using association rules.

We analyzed a dataset comprising student responses to geometric, linguistic, and algebraic questions by applying association rule mining techniques. The extracted rules were used to evaluate question similarity, revealing deeper insights into student performance and problem-solving strategies.

Our findings demonstrate significant interconnectedness between geometric, linguistic, and algebraic thinking, with implications for student performance measures and grading. Students' ability to solve problems in one domain often translated into enhanced performance in others, suggesting a shared set of cognitive resources and strategies. Association rules proved valuable for identifying nuanced relationships between question types and domains, providing a comprehensive perspective on student performance.

These results have important implications for educational practices, emphasising the need to consider the interconnectedness of learning domains when designing assessments and grading systems. By adopting a holistic approach to student evaluation, we can better support students' development of critical thinking and problem-solving skills across various domains, fostering deeper subject matter understanding and enhancing educational outcomes.

[1] *Corresponding Author: Bence Sipos*

bence.sipos@bme.edu.hu

1 INTRODUCTION

1.1 Problem

The evaluation of student performance has long been a central concern for educators and policymakers, particularly at the beginning of university studies when students face new academic challenges and may experience a higher risk of dropout. Accurate assessment of student performance is vital for several reasons, including providing targeted support to students who may be struggling, identifying talented individuals who could benefit from extra-curricular opportunities, and ultimately improving retention and academic success rates. In this context, our work examines the need to measure student performance early in their university careers and explores innovative methods to achieve this goal.

One of the primary reasons for assessing student performance at the onset of higher education is the high dropout rate observed in many countries, including Hungary. Early identification of students at risk of disengaging from their studies can facilitate timely interventions, helping to prevent attrition and promote academic achievement. By understanding the unique challenges faced by students in the transition to university life, educators can tailor support services and resources to maximize student success.

In Hungary, university admission scores are primarily based on high school final exam scores, which are not field-specific. This approach may not accurately predict a student's aptitude or preparedness for their chosen field of study, leading to potential mismatches between students' abilities and the demands of their academic programs. An improved method of assessing student performance that accounts for field-specific knowledge and skills could better align students with appropriate courses and facilitate smoother transitions into higher education.

1.2 Interconnectedness

The research question aims to explore how student test results can reveal the interrelationship between different areas of knowledge or skills. By providing layered responses during the evaluation of test results, a more precise understanding of the studied population can be obtained, and previously unknown hidden connections, such as those between mathematics and language, can be uncovered. This research question seeks to enhance the discovery of complex connections and relationships, contributing to a better understanding of student performance.

To address these issues, our study proposes a novel approach to evaluating student performance that takes into account the interconnectedness of learning domains. By identifying and leveraging associations between geometric, linguistic, and algebraic thinking, we aim to provide a more comprehensive understanding of student performance across multiple subject areas. The insights gained from this analysis can be used to design targeted interventions, such as offering extra-curricular classes for talented students or providing additional support for those who may be struggling.

In conclusion, the accurate and timely assessment of student performance at the beginning of university studies is crucial for supporting students' academic success and reducing dropout rates. By adopting a holistic approach that considers the interconnectedness of learning domains, we can better align students with appropriate courses, foster deeper subject matter understanding, and ultimately enhance educational outcomes.

.

2 METHODS

We applied association rules which are developed for the analysis of consumption patterns. To our knowledge, this method has not been used for the analysis of educational data. This allowed us to use a methodologically new tool to answer our research question.

2.1 Tests

Our research group focuses on creating entrance tests that accurately predict student performance and identify those needing extra support or talent development courses. By combining mathematical tests with language tasks, we can reduce learning bias-related distortion. A large student sample showed that language tasks help predict dropout rates and poor performance more precisely. Students excelling in both math and language tests tend to perform well in calculus subjects. Language tests, less dependent on grammatical knowledge, offer a more inclusive assessment of students' abilities.(Sipos et al. 2021)

The evaluation of geometric thinking was conducted using the van Hiele test (Senk 2022), while the assessment of linguistic and algebraic/reasoning abilities was performed through our custom-designed test. The van Hiele test consists of 25 questions, grouped in sets of five, whereas our test contains 14 questions related to mathematics and 31 linguistic/logic questions. (Olah e. al. 2019, Sipos snd Szilágyi 2022)

2.2 Association rules

To investigate the interconnectedness and correlation between questions in our test suite, we employed association rule mining, a widely recognized method for analyzing consumer behaviour. Association rule mining (Hipp 2000) is a powerful technique for discovering relationships and patterns among variables in large datasets. It has been particularly successful in market basket analysis, where it is used to identify items frequently purchased together. In our study, we applied this method to explore relationships between geometric, linguistic, and algebraic thinking questions.

Association rule mining relies on three primary measures: support, confidence, and lift. These measures provide valuable insights into the strength and relevance of the discovered relationships. Support refers to the proportion of the dataset in which an association rule is found to be true. A higher support value indicates that the rule occurs more frequently and is therefore more significant.

Confidence measures the likelihood that a particular association rule holds true. It is calculated as the ratio of the support of the entire rule to the support of its antecedent (i.e., the first part of the rule). A higher confidence value suggests a stronger relationship between the items in the rule. Lift is a metric that quantifies the improvement in prediction brought about by the association rule. It is calculated as the ratio of the confidence of the rule to the support of its consequent (i.e., the second part of the rule). A lift value greater than 1 indicates a positive association between the items in the rule, while a value less than 1 suggests a negative association.

In our analysis, we treated the solved questions as customer baskets. We applied association rule mining to these baskets to identify patterns and relationships between the different types of questions using the apriori algorithm provided in

‘arules’ package (Hahsler, 2005) version 1.7-6 in R (version 4.2.1). By uncovering these associations, we aimed to gain deeper insights into students' problem-solving strategies and performance across the geometric, linguistic, and algebraic thinking domains. The results of this analysis were used to refine the test suite, enhancing its efficiency and effectiveness in assessing student performance.
Association rules have limitations regarding large datasets and sparse data, which can affect the validity of results. This method primarily suits categorical or binary data, necessitating the conversion of continuous data into categorical by discretization.

3 DATA

The dataset used in this study was collected from 153 students who began their studies in 2020 in Mechatronics Engineering or Energy Engineering programs. Mechatronics Engineering requires the highest admission scores in Hungary among all engineering disciplines, while Energy Engineering also demands high admission scores. As a result, the two groups of students can be considered more talented than the average engineering student population.
Despite their high admission scores, the students in our sample exhibited diverse high school backgrounds. Some students had taken advanced-level mathematics courses in high school and completed advanced-level final exams, while others pursued advanced-level final exams in different subjects. This variation in prior knowledge has led to significant differences in mathematical proficiency among students, posing challenges in building a solid foundation in math for more engineering-oriented subjects.
Our dataset includes information on each student's performance in the geometric, linguistic, and algebraic thinking questions from our test suite. By analyzing this data using association rule mining, we aimed to identify patterns and relationships between the different types of questions and uncover insights into students' problem-solving strategies and performance across the three learning domains. This information can then be used to inform targeted interventions and support services, helping to bridge gaps in knowledge and ensure that all students can succeed in their engineering studies.

4 RESULTS

Upon analyzing the dataset, we observed that certain questions were solved by almost everyone, such as a question involving logarithmic equations or deciphering the meaning of a Latin word from a few Latin sayings containing the word. With over 94% of students successfully answering these questions, the information gain from these items was limited.

To focus on more informative associations, we narrowed our analysis to rules with low support (0.4) and high lift values. High lift values allowed us to identify relationships between similarly solved questions. Relying solely on correlation was insufficient due to the excessive noise (sometimes students make random mistakes), so the lift metric was employed to derive more meaningful insights.

It is important to note that the number of possible rules in our analysis was very high, while the number of observations (i.e., students) was relatively low. Consequently, we limited the maximum number of questions in one rule and used different support thresholds compared to those typically employed in other fields where the number of observations is much higher. By adjusting our analysis parameters to suit the unique characteristics of our dataset, we were able to uncover valuable information about students' problem-solving strategies and performance across the geometric, linguistic, and algebraic thinking domains.

4.1 Findings

The analysis of our dataset revealed several key findings about the relationships between different types of questions and the performance of students across geometric, linguistic, and algebraic thinking domains.

Questions from the same math subfield exhibited strong correlations: For instance, solving four of the last five van Hiele questions demonstrated a strong association with solving difficult geometry questions in our test. This suggests that students who perform well in one aspect of a math subfield tend to excel in other related problems within that subfield.

Connections between seemingly unconnected fields were identified: An interesting association was observed between solving the last van Hiele-level questions and finding synonyms in the linguistic portion of the test. This finding implies that there may be underlying cognitive strategies or abilities that are shared between geometric and linguistic thinking.

Advanced-level final exams and taking advanced math classes in high school not only led to better procedural math skills but also showed a strong positive correlation with several linguistic questions: Students with advanced math backgrounds performed better on tasks such as identifying the non-fitting word from a list or finding a word that could create a new complex word when combined with a list of words. This result suggests that there may be transferable skills or knowledge gained from advanced math coursework that also benefits linguistic performance.

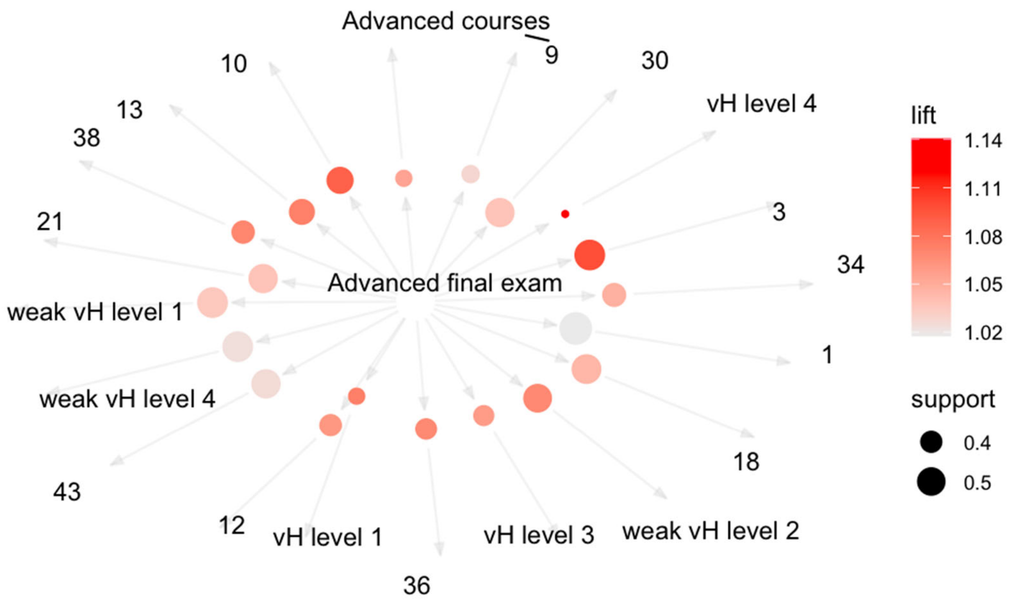

Fig. 1. The top 20 association rules with the highest lift value where the advanced level final exam is the antecedent (LHS).

Equations with real numbers and filling out missing chunks of text: Our analysis revealed a relationship between solving equations with real numbers and completing text-based tasks that require filling in missing portions. This finding indicates that students who are proficient in algebraic thinking may also possess strong linguistic skills or vice versa, further highlighting the interconnectedness of learning domains.

These results underscore the importance of considering the interconnectedness of learning domains when designing assessments and grading systems. By recognizing the complex relationships between different subject areas, educators can better support the development of critical thinking and problem-solving skills across various domains, ultimately enhancing educational outcomes for all students.

5 CONCLUSION

We aimed to employ association rules, commonly used in commerce and economics, to analyze educational data and map the relationship between students' knowledge in different areas. This approach allowed for a more sophisticated assessment of student knowledge.

Our findings demonstrate that association rules can effectively generate valuable insights into the relationships between questions in tests, shedding light on the interconnectedness of learning domains. However, there are several limitations to this approach. One significant challenge is the manual parameter tuning required for association rule mining, which can be labour-intensive and time-consuming. Additionally, the limited number of samples in our dataset restricts the confidence we can place in the discovered rules. A larger dataset would enable more robust and reliable results.

Despite these limitations, our study provides a novel approach to assessing exam data with potential applications in improving the testing process. By using the insights gained from association rule mining, educators can streamline examinations in two main ways:

Eliminating redundant questions: By identifying questions that are solved by almost everyone or removing highly correlated questions, the test can be shortened without sacrificing its ability to evaluate students' knowledge and skills effectively.

Implementing adaptive testing: Our findings can inform the development of an online exam system that recommends questions based on students' previous answers. This approach allows for the acquisition of more information about student performance within the same number of questions, optimizing the test-taking experience for both students and educators.

Future research could focus on expanding the dataset to include more students and diverse educational backgrounds, which would enhance the generalizability and reliability of the association rules. Additionally, further exploration of adaptive testing approaches and the development of algorithms to automate parameter tuning in association rule mining could lead to more efficient and effective assessment methods, ultimately benefiting both students and educators in the long run.

Another significant advantage of using association rule mining to analyze test data is its potential to differentiate more effectively between students. By identifying patterns and relationships between different types of questions and learning domains, educators can gain a deeper understanding of each student's strengths and weaknesses. This information can then be used to develop personalized learning plans, tailored interventions, and targeted support services to address individual needs and promote academic success.

Moreover, by incorporating adaptive testing strategies based on the insights gained from association rule mining, exams can be tailored to challenge and engage students at various ability levels. This approach not only provides a more accurate assessment of each student's performance but also fosters an inclusive learning environment that accommodates diverse learning needs.

In summary, utilizing association rule mining in conjunction with adaptive testing methods can help educators differentiate between students more effectively, ultimately enabling them to provide targeted support and personalized learning experiences that cater to individual strengths, weaknesses, and learning styles. This approach not only enhances academic outcomes but also promotes equity and inclusivity in education.

The association method is particularly well suited for analysing the results of complex tests in areas requiring a high degree of knowledge transfer, such as engineering education. For example, it can be used to investigate how knowledge of eigenvectors and eigenvalues in linear algebra is used to understand the stress tensor in material structure. Open-source software like R makes the method accessible to all. This not only allows us to quantify the effectiveness of learning and teaching but also to explore various previously unknown relationships.

6 ACKNOWLEDGMENTS

The research received funding from the MTA-ELTE Theory of Learning Mathematics Research Group.

REFERENCES

HAHSLER, Michael; GRÜN, Bettina; HORNIK, Kurt. arules-A computational environment for mining association rules and frequent item sets. Journal of statistical software, 2005, 14: 1-25.

HIPP, Jochen; GÜNTZER, Ulrich; NAKHAEIZADEH, Gholamreza. Algorithms for association rule mining—a general survey and comparison. ACM sigkdd explorations newsletter, 2000, 2.1: 58-64.

SENK, S. L., et al. The Van Hiele Geometry Test: History, Use, and Suggestions for Revisions. 2022.

OLÁH, Tibor; BEREZVAI, Szabolcs; SIPOS, Bence; SZILÁGYI, Brigitta. 2021. "Nyelvi–matematikai Kognitív Mérőteszt Alkalmazása a Felsőoktatásba Belépő Hallgatóknál." Szakképzés-Pedagógiai Tudományos Közlemények 1: 189–199.

SIPOS, Bence; OLÁH, Tibor; SZÉLES, Katalin; BALOG, Janka; SZILÁGYI, Brigitta, How to utilize test results effectively?. In: 49th SEFI Annual Conference: Blended Learning in Engineering Education: Challenging, Enlightening-and Lasting?. European Society for Engineering Education (SEFI), 2021. pp. 500-508.

SIPOS, Bence; SZILÁGYI, Brigitta, Investigation of Processing Test Results Based on Knowledge Similarity. In: Järvinen, Hannu-Matti; Silvestre, Santiago; Llorens, Ariadna; Nagy, Balàzs SEFI 2022 50th Annual Conference of The European Society for Engineering Education Proceedings: Towards a new future in engineering education, new scenarios that European alliances of tech universities open up, Barcelona, Universitat Politécnica de Catalunya (2022) pp. 710-719.