

Looking for Log-Based Traces of Mathematical Reasoning

Linoy ASOULIN^a, Arnon HERSHKOVITZ^{a*} & Michal TABACH^a

^a*School of Education, Tel Aviv University, Israel*

*arnonhe@tauex.tau.ac.il

Abstract: We explore the potential of log-based analysis to capture students' mathematical reasoning. Twelve secondary school students worked with a GeoGebra applet on identifying reflective symmetry, while their justifications and actions recorded. Ten types of reasoning were identified through qualitative analysis; two of which were operationalized via log data, however with only low-moderate alignment with students' explanations. Reasoning-independent logged data was not predictive of any of the reasoning types. Findings highlight the need for context-dependent, pedagogically informed log variables for assessing and supporting mathematical reasoning.

Keywords: Log analysis, mathematics education, classification task, mixed methods

1. Introduction

Mathematical reasoning is the ability to carry out logical and analytical thinking processes to get conclusions both deductively and inductively. Mathematical reasoning is focused on how people justify, infer, explain, and prove within mathematics. It is foundational to mathematical thinking and mathematical problem solving, hence promoting mathematical reasoning is considered key in mathematics education (Brodie, 2010).

Still, assessing mathematical reasoning remains a challenge, largely because communication is integral to reasoning. Students' explicit explanations must be evaluated, which raises methodological difficulties. Mostly, it was done by analyzing students' written or oral explanations, even when digital learning environments were used (Jensen & Skott, 2022). Learning Analytics may promote this field of inquiry, however it has been scarcely used (Rundquist et al., 2024). This is the gap that we bridge in this study.

Our objective is to test associations between log-based variables and reasoning, for enabling the assessment and support of mathematical reasoning in real-time. We set up the following research questions: 1) What types of mathematical reasoning arguments do students express when working on a digital classification task related to symmetry? 2) To what extent can these arguments be operationalized via the task's log files? 3) To what extent can reasoning-independent log-based variables be associated with these arguments?

2. Methodology

2.1 Research Participants and Research Process

We studied a convenient sample of twelve 8th-10th-grade students (ages 14-16 years old) from Israel. The first author had met each of the participants in a one-on-one session. After making sure they are familiar with the task interface and with the topic of reflective symmetry, they were asked to work on the main task (see next section) to its completion. Participants were asked to justify aloud each of their actions while they take them.

2.2 Research Tool and Data Collection

We used an applet integrated, designed and developed using GeoGebra (Noster et al., 2024). The applet presents users with seven quadrilaterals, which they are asked to classify:

square, tilted square, parallelogram, 90-degree rotated parallelogram, trapezoid, kite, and an irregular shape (see Fig. 1); classification is based on the existence of a line of symmetry, and is done by dragging the shapes into one of two regions (“One of more lines of symmetry”, or “No lines of symmetry”). Immediate feedback is available in the form of an updated cumulative count of correct and incorrect classifications. Users can keep dragging objects from anywhere to anywhere on the screen. While working on the task, log files of the applet were collected, as well as screen recording and participants’ justifications.

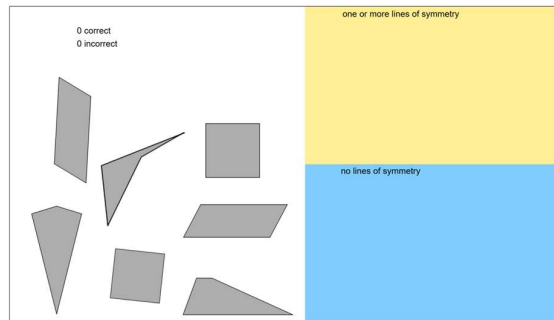


Figure 1. The GeoGebra Applet Used in this Study.

2.3 Data Analysis

Annotated screen recordings and transcripts of participants’ recordings were aligned. Unit of analysis was a classification attempt, with 101 analyzed drags. Using this data, we identified participants’ reasoning (RQ1) using a thematic analysis, via an iterative process with frequent discussion between the authors until agreement. The final scheme included ten categories of reasoning; a single drag could have been assigned to multiple categories.

Log files were used to extract two types of variables at the drag level. First, reasoning-derived variables, which were operationalized based on the participants’ reasoning scheme (RQ2). Second, reasoning-independent variables (RQ3) that were calculated regardless of the coding: 1) Distance from a Straight Dragging Line – based on coordinates of mouse movement which were logged in 0.02 second-intervals; 2) Time from Previous Drag – calculated as the difference in seconds between first timestamp of current drag and last timestamp of previous drag, normalized using Z-score regarding each participants’ drag to account for variability in thinking- and response times; 3) Drag Duration calculated as the difference in seconds between last and first timestamps of current drag, normalized in the same way as the previous variable. Values of both variable types were eventually aligned with the qualitative data to allow for statistical analysis.

3. Findings

3.1 Mathematical Reasoning: Justification of Symmetry-Based Classification (RQ1)

Participants justified 94 drags, and we found ten types of mathematical reasoning (see Table 1). Note that a student’s reasoning is not dependent on the classification outcome, that is, following a certain reasoning may lead to either correct or incorrect result. Importantly, some drags were coded to multiple categories – over a quarter of them (25 drags) were coded to two categories, and about 12% (11 drags) were coded to three categories; one drag was coded to four categories.

3.2 Capturing Mathematical Reasoning in Log Files (RQ2)

We found only two reasoning categories that could be potentially identified via the log files. The analysis in this section is based on the whole dataset, i.e., N=101 drags, including those who were not manually coded.

Table 1. *Types of Reasoning for Justifying Symmetry-Related Classification (N=94 Drags); highlighted with a grey background are reasoning types that we operationalized via log files*

Reasoning Type	Explanation	Example (Participant Code, Referred Shape)	#
Matching Two Parts of the Shape	Partitioning the object to two, based on size, vertices, or edges.	"The lower edge is in an angle, also the upper edge, but on different sides, so they'll cover each other" (S6, rotated parallelogram)	32
Using a Line	Searching for horizontal, vertical, or diagonal line of symmetry	"Here you can draw a height, it will create two equal triangles" (S2, kite)	30
Folding or Mirroring	Imagining a folding of the shape or mirroring it on itself	"If I fold it in the middle, this vertex will get to that one, and this edge will get to that one, and it will be symmetrical" (S12, kite)	22
Classifying Objects by Similarity	Relying on a similar object as a reference	"Because it's similar to that shape" (S7, rotated parallelogram)	22
Intuitiveness	"Knowing" the answer intuitively	"The brain gets it immediately" (S5, square)	12
Comparing Edges, Angles	Using information on edges, angles	"Because there is this small line, and here you don't have it" (S11, trapezoid)	11
Using Feedback for Incorrectness	Acting upon feedback when count of incorrect classification increases	"I understood [from the feedback] why it's a mistake, now I looked at it accordingly" (S1, rotated parallelogram)	8
Prior Knowledge	Using prior knowledge about symmetry of shapes	"It looks like a kind of a kite, and I know that kite is symmetric" (S6, kite)	3
Weird Shape	Referring to shape's irregularity as an indicator for non-symmetry	"There's not at all, like, a shape here" (S5, irregular)	2
Tilted Shape	Arguing that a tilted shape is not symmetrical	"This one is crooked, so it's not symmetrical" (S8, tilted square)	2

First, *Classifying Objects by Similarity*. Our operationalization: If a given drag involved a similar shape to the one dragged immediately previously (square and tilted square, or parallelogram and 90-degree rotated parallelogram) and if both were classified alike, it is an indication of using that reasoning. Calculating agreement between the log-based metric and students' evidence resulted with a small-moderate Cohen's Kappa of 0.29. Re-examining the data, we could explain three misalignments: 1) A student who used this reasoning mentioned it regarding the classification of the first of two similar objects, while the operationalization was based on dragging the second object; 2) A student dragged square and tilted square a total of eight times consecutively between regions, but only once mentioned using this reasoning; 3) A student justified their same-classification of similar objects using this argument, but mistakenly dragged the second shape to a different region than the first one. Correcting for these, we got a moderate-high Kappa of 0.61.

The second reasoning category we can operationalize is *Using Feedback for Incorrectness*. We did so as follows: If a given drag involved the same shape as in the immediately previous drag, and if the previous drag was incorrectly classified and the current is correctly classified – it is an indication of using the feedback. Here again, we got a small-moderate Kappa of 0.30. Re-examining the data, we noticed that in two cases, participants who claimed using this reasoning did not immediately correct their incorrect classification in the next step after the feedback indicated on a mistake (see discussion).

3.3 Reasoning-Derived Log-Based Evidence Vs. Student' Argumentation

For keeping a reasonable statistical power, we only referred to those categories that were evident in at least twenty cases, i.e., *Matching Two Parts of the Shape*, *Using a Line*, *Folding or Mirroring*, and *Classifying Objects by Similarity*. For each of these—coded for each drag as evident or not—we built a multivariate logistic regression model, based on the three reasoning-independent log-based variable. Neither of the models was significant, which means that the use of these argumentations cannot be predicted based on the variables we defined. See a summary of the models in Table 2.

Table 2. *Logistic Regression Models for Matching Two Parts of the Shape, Using a Line, Folding or Mirroring, and Classifying Objects by Similarity (df=87)*

Variable	<i>Matching Two Parts of the Shape</i>		<i>Using a Line</i>		<i>Folding or Mirroring</i>		<i>Classifying Objects by Similarity</i>	
	Estimate (SE)	Wald Test	Estimate (SE)	Wald Test	Estimate (SE)	Wald Test	Estimate (SE)	Wald Test
Intercept	-0.60 (0.24)	6.4 p<0.05	-0.78 (0.26)	9.1 p<0.01	-1.40 (0.27)	27.0 p<0.001	-1.25 (0.26)	23.0 p<0.001
Distance from Straight Dragging Line	-0.01 (0.02)	0.26 p=0.61	-0.04 (0.05)	0.46 p=0.50	0.00 (0.00)	2.0 p=0.16	0.00 (0.00)	0.99 p=0.32
Time from Previous Drag	-0.34 (0.27)	1.66 p=0.20	0.05 (0.26)	0.04 p=0.84	-0.15 (0.31)	0.25 p=0.62	0.25 (0.27)	0.83 p=0.36
Drag Duration	0.17 (0.27)	0.40 p=0.53	-0.05 (0.29)	0.03 p=0.89	-0.29 (0.35)	0.70 p=0.40	0.22 (0.29)	0.55 p=0.46
Nagelkerke R ²	0.08, p=0.14		0.06, p=0.29		0.05, p=0.36		0.05, p=0.40	

4. Discussion

In this exploratory study, we looked for log-based evidence for students' mathematical reasoning. We found ten reasoning types which align with the literature (Kollosche, 2021), and then looked for evidence for them in logged traces, by using either reasoning-derived or reasoning-independent variables; overall, the former is a more powerful approach than the latter. Mathematical reasoning is highly sensitive to the specific content involved. Therefore, thinking of a "global" Learning Analytics solution for the "mathematical reasoning problem" is an implausible idea. Rather, Learning Analytics must adapt to the learning context in which it operates (Shibani et al., 2019).

The alignment between students' explanations and their log-based operationalization was only moderate. In some cases, this derived from a timing discrepancy: While our operationalization refers to actions that are expected to co-occur with reasoning, participants often did not act in this simultaneous manner. It is recommended to account for the probability of such a delayed action. The moderate association can be also explained by participants not always expressing their full set of argumentations. Finally, multiple reasoning co-occurring had obscured our ability to predict reasoning; from the vast Knowledge Tracing literature, we know that multi-skill formulation increases model complexity and computational requirements (Krivich et al., 2025).

While limited in by the fact that we could only operationalize two of the ten reasoning types—which may be addressed by using eye- or mouse-movement tracking—and in sample size, this study emphasizes the importance of understanding the problem, both content- and interaction-wise, when wishing to study it via log files.

References

- Brodie, K. (2010). Teaching mathematical reasoning: A challenging task. In K. Brodie (Ed.), *Teaching Mathematical Reasoning in Secondary School Classrooms* (pp. 7–22). Springer.
- Jensen, E. O., & Skott, C. K. (2022). How can the use of digital games in mathematics education promote students' mathematical reasoning? A qualitative systematic review. *Digital Experiences in Mathematics Education*, 8(2), 183–212.
- Kollosche, D. (2021). Styles of reasoning for mathematics education. *Educational Studies in Mathematics*, 107(3), 471–486.
- Krivich, E., Hooshyar, D., Šír, G., Yang, Y., Bauters, M., Hämäläinen, R., & Kärkkäinen, T. (2025). A systematic review of deep knowledge tracing (2015-2025): Toward responsible AI for education. *Preprints*, 2025101845.
- Noster, N., Hershkovitz, A., Tabach, M., & Siller, H.-S. (2024). Classification tasks as basis for assessment - Family resemblance as principle for choosing mathematical objects. *Proceeding of FAME - Feedback & Assessment in Mathematics Education*.
- Rundquist, R., Holmberg, K., Rack, J., Mohseni, Z., & Masiello, I. (2024). Use of learning analytics in k–12 mathematics education. *Journal of Learning Analytics*, 11(3), 174–191.
- Shibani, A., Knight, S., & Shum, S. B. (2019). Contextualizable learning analytics design. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, 210–219.