ZDM – Mathematics Education (2024) 56:727–739
https://doi.org/10.1007/s11858-024-01551-5

ORIGINAL PAPER

Learning analytics in mathematics education: the case of feedback use in a digital classification task on reflective symmetry

Arnon Hershkovitz[1] · Norbert Noster[2] · Hans-Stefan Siller[2] · Michal Tabach[1]

Accepted: 18 January 2024 / Published online: 3 March 2024

Abstract

Learning Analytics is concerned with the use of data collected in educational settings to support learning processes. We take a Learning Analytics approach to study the use of immediate feedback in digital classification tasks in mathematics. Feedback serves as an opportunity for learning, however its mere existence does not guarantee its use and effectiveness, as what matters is how learners interact with it. Therefore, our research questions are focused on that interaction. The data consisted of 266 object movements for classifying polygons, and 524 shape movements for classifying traffic signs, under the topic of symmetry. Participants included 29 elementary school students (9–12 years old) from Israel and Germany. Analyzing students' success, feedback use, and the associations between them, we demonstrate how not acting upon feedback is negatively associated with success, and how this undesired behavior slightly reduces along the learning process.

Keywords Digital classification task · Elementary school · Learning analytics

1 Introduction

Digital technologies increase opportunities for learning and teaching, as well as for educational research. In this context, Learning Analytics—an area of research and practice in education that takes a computational approach to understand and improve learning processes—enables detecting patterns of usage in digital learning environments (Chatti et al., 2014). This is what we intend with our study, in the context of feedback use in mathematics education.

While solving digital tasks, automatic and immediate feedback may serve as an opportunity for learners to improve their knowledge and skills. A recent literature review suggests that in most cases, automatic feedback boosts student performance (Cavalcanti et al., 2021). Nevertheless, the mere existence of feedback does not imply on its use. Whether feedback is used is mostly up to the learner (Narciss, 2013; Winstone et al., 2016). Yet, research on students' actual use of feedback throughout the learning process is only in its infancy. Recently, Iraj et al. (2021) stated that to the best of their knowledge, theirs was "the first study to use an empirical methodology to examine student engagement with technology-mediated feedback" (p. 112). Although we cannot make such a strong claim, we do share the feeling that not much has been done in this field so far. For example, Bruder et al. (2021) mentioned some studies examining feedback in mathematics education, for example diagnostic tests with automated individualized feedback (Roder, 2020) or digital diagnostic environments for identifying fault patterns (Nitsch, 2015). Still, there is a gap in researching student engagement with technology-mediated feedback.

Here we take a step forward in bridging this gap by using Learning Analytics, specifically analyzing log files drawn from a digital learning environment; these files continuously and automatically document student actions within the environment. We do so in the context of two-choice classification tasks, which provide a convenient opportunity for exploring feedback use. In such tasks, which require students to classify objects based on a binary criterion (e.g., has/

✉ Arnon Hershkovitz
arnonhe@tauex.tau.ac.il

Norbert Noster
norbert.noster@mathematik.uni-wuerzburg.de

Hans-Stefan Siller
hans-stefan.siller@mathematik.uni-wuerzburg.de

Michal Tabach
tabachm@tauex.tau.ac.il

[1] Tel Aviv University, Tel Aviv, Israel

[2] Julius-Maximilians-Universität Würzburg, Würzburg, Germany

does not have, or has A/has B), simple feedback that indicates whether the classification is correct implies the correct answer even when a classification was wrong. Therefore, by studying students' behavior while engaged with digital two-choice classification tasks that offer immediate feedback on correctness or incorrectness, we can examine how students use feedback. This is indeed the goal of the current research.

2 Literature review

2.1 Feedback in digital learning environments

Feedback is key to learning. Yet, as feedback can take multiple forms, it is not assumed a priori to have a positive impact on learning. In reviewing meta-analyses of the power of feedback, Hattie and Timperley (2007) stressed that the most effective feedback involves information about a task and how to perform it effectively, including cues or reinforcement. These findings were replicated by an updated meta-analysis (Wisniewski et al., 2020) that highlights the importance of providing information beneficial to learners. Beneficial feedback should be task-related, help students reject erroneous hypotheses, and direct them in searching and strategizing (Hattie & Timperley, 2007). This already hints towards the important role of the learner in the feedback use. Indeed, Narciss (2013) emphasized that learners may be affected by their goals, competencies and motivation even before they begin processing feedback information. They may subsequently undergo a self-controlled process of processing and learning that may lead to adjustments at their cognitive, meta-cognitive and motivational levels. For feedback to be effective, learners should receive it proactively through a self-informed, meta-cognitive processes (Winstone et al., 2016). This is supported by a recent review of feedback models that highlights the importance of the learner (Panadero & Lipnevich, 2022).

Studies have shown that students' preferences regarding feedback type in digital learning environments (e.g., no feedback, correct answer, explanation alone, or answer accompanied by an explanation) vary greatly. Moreover, a high proportion of learners pay very little or no attention to feedback (Schimmel & Schimmel, 1988; Timmers & Veldkamp, 2011). That computer-based elaborated feedback was found ineffective, led Golke et al. (2015) to draw the reasonable conclusion that this was a result of "the feedback was ignored rather than that it contained useless information" (p. 133). Therefore, feedback quality itself is not sufficient to determine its usefulness for learning, and we must further investigate whether learners notice feedback and act upon it.

To study learners' engagement with feedback, we choose a digital environment that automatically provides immediate feedback to the learners during the process of solving a task. This way we have multiple instances in which we can observe possible (non-)reactions to feedback. Furthermore, in the case of a binary classification which we study, feedback on correctness may drive students to immediately correct incorrect responses, hence it serves the purpose to study learners' engagement with feedback.

2.2 Learning analytics in mathematics education

Along with the ever-growing use of digital learning environments, the affordances of using log files for educational research has been emphasized. Log files automatically and continuously document student actions within the system. Most commonly, these files document learner actions, characterized in three dimensions: the action taker (who?), the action itself (what?), and the action time (when?). Analyzing this data, it is possible to calculate a wide range of measurements of the learning process, with the overall goal of improving learning (Duval & Verbert, 2012).

Log-based educational studies lie under the umbrella of Learning Analytics, which is most commonly defined 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" (Ferguson, 2012a, p. 305). Within this wide area of inquiry, modeling student behavior in digital learning environment has been one of the key goals (Papamitsiou & Economides, 2014), and this is indeed the angle we take here. Note that traditionally, as Ferguson (2012b) put it, "learning analytics and its related fields are strongly focused on higher education", but recently its importance to primary and secondary education has been stressed (Baker & Koedinger, 2018). The current study is another step in that direction, which continues our line of research in implementing log-based studies in mathematics education (Ben-Haim et al., 2019; Hershkovitz et al., 2019, 2021, 2023).

Analyzing student behavior in mathematics education using learning analytics has raised insights that are important for learners, educators, content developers, and policymakers. One of the most prominent examples in that sense is the development and study of ASSISTments, a popular online tutor for mathematics, of which log-based studies have demonstrated the existence of different patterns of student behavior (Adjei et al., 2017). Many other digital learning environments for mathematics have been used in log-based studies to explore learners' behavior and we will mention just a few. Data that was collected using the Battleship Numberland educational game was used to learn about number line estimation skills among fourth- to sixth-grade students, as well as about the associations between level of difficulty and engagement (Lomas et al., 2013). Data logged from MATHia, an online software for middle-school mathematics, was analyzed for studying content personalization based on students' personal interests, while also looking for associations between students' success and personal interests (Fancsali & Ritter, 2014). Data from other intelligent tutoring systems for mathematics was used to explore patterns of behavior in the system—e.g., help-seeking, multiple attempts—and their links with achievements and learning (Velasquez et al., 2014; Xie et al., 2017). Logged data from an interactive digital geometry task, where fifth- and sixth-grade students were asked to construct, measure and transform geometrical objects, was used to portray patterns of problem solving (Leiba, 2010). In the context of studying feedback in mathematics, log files were used in different settings to measure students' learning while getting different types of feedback (de Kock & Harskamp, 2016; Gal & Hershkovitz, 2019).

Key to learning analytics is the concept of operationalization, that is, the translation of a theoretical construct to a log-driven computational mechanism. This process requires a deep understanding of both the theory of the construct in question and the nature and meaning of the stored data. Obviously, data from log files which is limited to student actions within a learning environment—contrary, for example, to data that is originated in students' eye movement and gazing—cannot inform us on viewing feedback. However, we may infer from it about acting upon feedback by knowing that feedback was given and examining the following action; this is the approach we take.

Regarding the operationalization of learner behavior, we should emphasize two methodological considerations. First, it is often easier to identify cases of not demonstrating a behavior rather than of demonstrating it, as has been common in log-based studies (Baker et al., 2004; Hershkovitz et al., 2013; Wixon et al., 2012). Second, learning periods are often split into sub-periods for tracking a learning behavior along the learning process; sub-periods vary based on the study context, and could refer to, e.g., quantiles (Hershkovitz & Nachmias, 2011), days or weeks (Colthorpe et al., 2015), or practice opportunities (Baker et al., 2013). Our approach follows these considerations, as we operationalize the absence of acting upon feedback, and for some analyses we split the period of using each task to two sub-periods.

3 Research questions

The presence of immediate feedback concerning the correction of an object classification attempt leads us to three research questions (RQs). First, we address learners' overall success in tasks, as learners could complete the task with full success if they pay close attention to the feedback, understand it, and act upon it. Therefore **RQ1a:** *To what extent do students succeed in a digital task that presents them with immediate feedback about object classification?;* **RQ1b:** *How does success change between tasks?*

Second, we address the extent to which students demonstrate a behavior of not acting upon feedback, and how this behavior changes along the learning process. Therefore, **RQ2a:** *To what extent do students not act upon feedback in a digital classification task that presents them with immediate feedback?*; **RQ2b:** *To what extent does behavior of not acting upon feedback change within and between tasks?*

Finally, we address the associations between not acting upon feedback and success, in order to understand the potential implications of this behavior. Therefore, **RQ3a:** *What are the associations between not acting upon feedback and success?*; **RQ3b:** *What are the associations between changes in not acting upon feedback within tasks and success?*

4 Methodology

4.1 Research population

Our population consists of a convenient sample of $N=29$ elementary school students (17 boys and 12 girls) from 4 to 6th grades, i.e., 9–12 years old ($M=10$, $SD=0.9$), from both Israel and Germany. In Israel $n_1=12$ students were recruited through personal networks of the research team. In Germany, a 4th-grade classroom was recruited with $n_2=17$ participants. In both countries, the sample contained students from different skill levels. We are aware of a statistically significant difference in age between the two country-based groups (Mann–Whitney's W-value = 38, at $p<0.01$, and with Rank-Biserial Correlation of 0.63).

In order to make sure that data from the two populations could be merged, we compared averages of all the research variables, using an independent samples Mann Whitney test. None of the variables showed significant difference between the countries. Also, there were no gender differences between the country-based groups, with $\chi^2(1)=2.26$, at $p=0.13$. Therefore, we treat the whole population as one group.

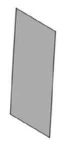
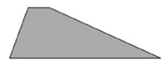
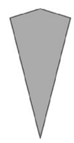
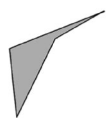
4.2 Research tool and a priori analysis

Reflective symmetry, the mathematical topic addressed here, was chosen because it is commonly taught from grade 3 at primary school on in Germany and Israel, and due to its strong visual presence, which makes it less language dependent. We focus on reflective symmetry of plane shapes, specifically on the recognition of lines of symmetry. Mathematical textbooks in both Germany (3rd-grade) and Israel (4th-grade) introduce students to reflective symmetry in a very similar way, by the following activity: Fold a piece of paper into two parts, and cut out a shape close to the fold line, without cutting that line; open the folded paper—the resulting shape has two congruent halves, it is a symmetric figure, and the folding line is the axis of symmetry (Kolbinger et al., 2006, p. 90; Rosenthal et al., 2010, p. 143). In the classification tasks, students were asked to classify objects based on the number of lines of symmetry. For that, two applets were designed and developed using GeoGebra (see Fig. 1). We chose GeoGebra as it allowed for flexible tasks design, and logging students' actions.

4.2.1 Research tool

The first applet—polygons task—presented seven quadrilaterals (Fig. 1, left), which the students were asked to classify based on the existence of at least one line of symmetry or none; the second applet—traffic task—presented ten traffic signs (Fig. 1, right), which the students were asked to classify based on having either a single line of symmetry or multiple lines. The objects used in these tasks are presented in Tables 1 and 2. In each task, classification was done by dragging objects into one of two regions; students were able to re-drag an object from one region to another, including to and from the pool region, at any point and as many times they wished. We ran the applet on either tablets or touch-screen laptops.

Table 1 The objects that were incorporated within the polygons task

Image	Name	Has/has no line(s) of symmetry	Intuitiveness
	Square	Has	Intuitive
	Tilted square	Has	Non-intuitive
	Parallelogram	None	Non-intuitive
	Rotated parallelogram	None	Non-intuitive
	Trapezoid	None	Intuitive
	Kite	Has	Intuitive
	Irregular	None	Intuitive

4.2.2 A priori analysis

4.2.2.1 Designing the polygons task While designing this task, we included both examples and non-examples as well as intuitive and non-intuitive figures (Tsamir et al., 2008). In the polygons task, the square and the kite (see Table 1) can be considered as intuitive examples (of having at least one line of symmetry) due to their typical representation. While the two parallelograms are presented in a typical shape (with the Rotated Parallelogram in an unusual orientation), they are non-intuitive non-examples, because they have a rotational symmetric property but not a reflective symmetry property; these can easily be mixed up and lead to a wrong classification. Since students seem to struggle with inclined lines of symmetry (e.g., Götz & Gasteiger, 2022; Seah & Horne, 2019) a rotated version of the square was added. As inclined lines of symmetry are considered non-intuitive, we added two intuitive non-examples: a trapezoid and an irregular concave quadrilateral.

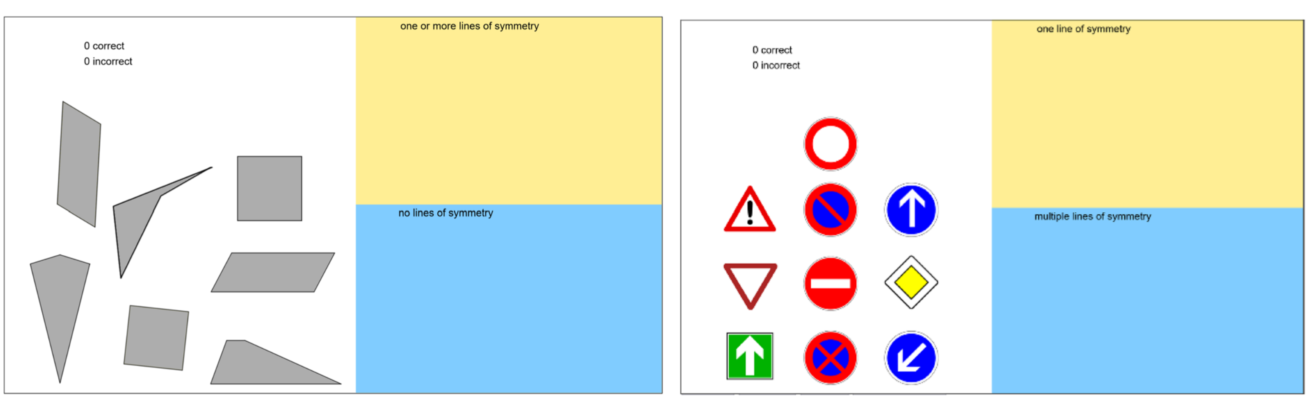


Fig. 1 The applets used in this study: classifying quadrilaterals by the existence/non-existence of at least one reflective symmetry line (left), and classifying traffic signs based on the existence of a single or multiple reflective symmetry lines

Table 2 The objects that were incorporated within the traffic task

Image	Name	Single/multiple lines of symmetry
	Priority road	Multiple
	Proceed straight	Single
	Road closed	Multiple
	Warning	Single
	Yield	Multiple
	No parking	Multiple
	Follow this way	Single
	No entrance	Multiple
	No stopping	Multiple
	Pass this side	Single

4.2.2.2 Designing the traffic task In the traffic task, we chose shapes that students are most likely to be familiar with in their everyday life, while at the same time not being necessarily discussed in school. As traffic signs may come with at least one line of symmetry, we chose to ask students to categorize them as having one or multiple lines of symmetry.

Signs may consist exclusively of *basic geometric shapes* such as triangles, circles, squares both as outer and inner shapes (e.g., Priority Road, Yield, No Entrance), as opposed to shapes such as arrows or exclamation marks. In these latter instances, the *inner shape* differs from the *outer shape* (e.g., Proceed Straight, Warning, No Parking). As mentioned above, the *orientation of the line of symmetry* (inclined vs. horizontal/vertical) is likely to be relevant for the difficulty level of their classification. Another factor is the number of lines of symmetry which varies from one (e.g., Warning) to infinite (e.g., Road Closed).

Since traffic signs are not often discussed regarding their geometric (and with that their symmetric) properties, classifying them as (non-)intuitive is complex. It can be assumed that an interplay between the previously mentioned characteristics (basic geometric shapes, orientation, and number of lines of symmetry) in a single object will affect learners' classification process, making it more difficult to predict their behavior.

4.2.2.3 Designing the feedback mechanism For the two tasks, we distinguish between task and subtask level, as decomposing a task or problem into subtask is a valuable heuristic (Polya, 1981). On the *subtask level,* each object classification is considered an individual (sub-)task, of making the correct binary choice. On the *task level* each applet is considered as one task, that is, to classify all objects; as students are presented with all objects at once, they should implement strategies or heuristics for successfully completing their task, and may be impacted, while solving it, on their previous actions and what they had learned from them.

Furthermore, we differentiate between surface and deep behaviors. On the *deep behavior,* objects are analyzed by the student and classified based on the (assumed) properties. A classification of a square as having multiple lines of symmetry goes along with the hypothesis that it owns multiple lines of symmetry. However, the classification of objects in these tasks can be handled merely based on the provided immediate feedback, moving an object from a certain location without any (reasonable) hypothesis to another, which we consider to be *surface behavior*. Note that we do not claim to measure understanding, but rather only observable actions.

These distinctions are relevant for the choice of our feedback design. Considering the task level, feedback should be provided after all objects have been classified. On subtask level feedback is necessary after each classification. Using this task as opportunity for students to learn, feedback on subtask level seems desirable (Drullion, 2019; Frey, 2022; Kuikka et al., 2016; McGuire et al., 2017) as it allows for drawing inferences for further classifications from (in-)correct classifications. Therefore, a design was chosen to be given at the subtask level. Yet, it should not interfere with any thought process while working on deep behavior. This is why feedback is provided in the form of an updated cumulative count of correct and incorrect classifications. It is rather low in attracting attention, while allowing for checking the correctness of all currently classified objects.

This feedback design also comes with a tradeoff, as it allows for students to work surface behavior and achieve correct classification, as explained above. Regardless of whether learners work on surface or deep behavior, it is expected that they correct any incorrect classifications immediately. This is as long as they pay attention to the feedback and treat classification of single objects as a subtask. Cases in which students do not immediately correct themselves may indicate that students either do not notice feedback or at least not react upon it.

4.3 Data collection

Data collection took place in March 2022. Members of the research team met with each of the participants individually. In Israel, meetings took place in the students' homes, after getting approval from their parents; in Germany, meetings took place in school, after getting approval from their parents and school team. Research was approved by the Institutional Review Board in both institutions.

First, participants were asked if they were familiar with the concept of reflective symmetry. Then, the researcher used another task, designed similarly to the two that we used and without any relation to symmetry, to ensure that the participant became familiar with the task interface and with the feedback mechanism. Finally, the researcher presented the participant with the symmetry applets, and made sure the instructions were clear. The participants were then allowed to use the applets by themselves until they stated that they were done. Each such meeting lasted for a few minutes.

4.4 Data preparation

Our data included data at the level of object-movement, that is, dragging and dropping an object from one place on the screen to another place, along with the correctness of the classification. Overall, for the polygons task we had 266 actions with the number of actions per participant ranging between 7 and 21 (M = 9.2, SD = 3.0); for the traffic task, we had an overall of 524 actions, ranging per participant between 10 and 69 (M = 18.1, SD = 13.9). This data served as the basis for our analysis.

4.5 Data analysis

Analysis was done separately for each applet. Due to the relatively small population size, we used non-parametric statistical tests, that is, testing for differences between paired samples using Wilcoxon's Signed-Rank test, and for correlations using Spearman's rho. Analyses were conducted in JASP 0.17.2.

4.6 Research variables

4.6.1 Not acting upon feedback (NAUF) [0–1]

We cautiously looked for cases in which noticing feedback and acting upon it would have resulted in a correct classification, but this did not happen. We identified two such scenarios: (1) When the most recently dragged object was correctly classified and in the immediate next step it was dragged out of its classification area; (2) When the most recently dragged object was moved to its incorrect destination, and in the immediate next step it was kept there. In other cases, we feel that inferring about not acting upon feedback is too interpretative. We first mapped all relevant steps—not including last action of each student—as either not acting upon feedback (1) or not (0). Then, we calculated a student-level NAUF by taking an average of the action-level values across their actions.

4.6.2 NAUF first/second half [0–1]

This variable accounts for changes in NAUF along the solution process referring to the two halves of a student's solution process (regarding number of steps). We calculated for each student the relative amount of NAUF for the first and second halves of their solution process.

4.6.3 NAUF trend [0–1]

This variable is a proxy for changes in NAUF along the solution process in yet another way, referring to the linear trend of cumulative NAUF values along the solution process. This is calculated as follows: We considered a function where the step number represents the independent variable and the cumulative NAUF score as the dependent variable. Then, we calculated the slope of the linear trend line of this function. A positive value denotes an overall increase in NAUF instances, and a value close to zero denotes an overall steady behavior.

4.6.4 Final correct classifications [0–1]

For each student, we counted the number of correct classifications when they stated they had completed the task. The maximum value for this variable is 7 for the Polygons applet, and 10 for the Traffic applets, therefore we normalized this variable by the total number of objects in each applet, for being able to compare between the tasks.

4.6.5 Overall success [0–1]

For each student, we divided the number of correct classifications upon completing the task by the total number of actions it took to complete the task, to account for both correctness and number of attempts required to achieve the solution. The closer the value of this variable is to 1—the better the success is.

5 Findings

This section is organized based on our RQs. In each subsection, we present findings for each task separately.

5.1 Success (RQ1)

We first ran descriptive statistics of the two success variables, then compared their means between the two tasks. Results are presented in Table 3 and described next.

In the polygons task, o*verall success* took a medium–high mean (0.79), and *final correct classifications* took a very high average of 0.96; there were 7 participants (24%) who finished the task without all objects being classified correctly. In the traffic task, *overall success* took a medium–high mean (0.70), and *final correct classifications* took a very high average of 0.98; there were 5 participants who finished the task without all objects being classified correctly.

The differences in o*verall success* and *final correct classifications* between tasks were not significant. The difference between the tasks in the distribution of completing the task with full success (yes/no) was also non-significant. Note that in a previous analysis of the same data drawn from the polygons task, we observed that non-intuitive quadrilaterals were more difficult to successfully classify than intuitive ones; no significant difference was observed for example vs. non-example quadrilaterals (Noster et al., 2022b). As for the traffic task, intuitiveness was not a priori assumed, and we could not draw exclusive conclusions regarding difficulty of examples vs. non-examples (Noster et al., 2022a).

5.2 Not acting upon feedback (RQ2)

We first report on step-level analysis of NAUF, and then on student-level analysis, so we could get a full understanding of this behavior. Results are listed in Table 4 and are described next.

5.2.1 Step-level analysis

Descriptive statistics of the research variables is reported before presenting the comparison between the tasks.

In the polygons task, there were overall 34 cases of *not acting upon feedback*, which accounted for 14% of all shape movements (out of 237 relevant movements, i.e., not counting last movement for each student). Interestingly, there was a difference in percentages between the different types of not acting upon feedback. For the cases where an object was classified correctly, only 2 of the 174 cases (1%) resulted in not acting upon feedback. However, for the other scenario, the rate was much higher: 32 of the 50 cases where an object was incorrectly classified (64%). Furthermore, of the cases when dragging non-intuitive quadrilaterals, 20% resulted in NAUF (25 of 122), while only 8% of the cases when dragging intuitive objects resulted in NAUF (9 of 115). The rates of NAUF is the cases of classifying example vs. non-example objects were similar to each other, 13% (15 of 112) and 12% (19 of 154), respectively.

Table 4 Summary of the findings for student-level NAUF

Variable	Polygons M (SD)	Traffic M (SD)	Difference
Overall NAUF	0.12 (0.13)	0.09 (0.09)	$z = 0.88$ $P = 0.39$
NAUF first half	0.11 (0.16)	0.13 (0.14)	$z = 0.06$ $p = 0.97$
NAUF second half	0.14 (0.23)	0.06 (0.09)	$z = 1.60$ $p = 0.12$
Within-task difference between halves	$z = 0.31$ $p = 0.78$	$z = 2.24$*	
NAUF trend	0.14 (0.15)	0.08 (0.08)	$z = 1.64$ $p = 0.11$

*$p < 0.05$

In the traffic task, we had a slightly lower overall rate of *not acting upon feedback*, with 57 of 495 relevant movements (12%). For the cases where an object was classified correctly, only 5 of the 324 cases (2%) resulted in not acting upon feedback. For the cases where an object was classified incorrectly, 52 of the 153 cases (34%) resulted in not acting upon feedback. Finally, we look at NAUF instances from the perspective of dragged object's symmetry. We found that of the 301 movements of objects that had multiple lines of symmetry, 38 resulted in NAUF (13%), slightly above the rate regarding objects that had a single line of symmetry (10%, 19 of 194).

Overall, *not acting upon feedback* rate was not significantly differing between the two tasks, with $\chi^2 = 1.18$, at $p = 0.28$. In the scenario of *not acting upon feedback* that follows a correct classification, there was no significant difference between the two tasks, with $\chi^2 = 0.13$, at $p = 0.72$; the change in the second type of *not acting upon*

Table 3 Summary of the findings related to success

Variable	Polygons	Traffic	Difference
Overall Success	0.79 (0.19)	0.70 (0.24)	$Z = 1.37$ $p = 0.18$
Final Correct Classifications	0.96 (0.08)	0.98 (0.05)	$Z = 1.38$ $p = 0.18$
# Participants with no perfect score	7 / 29 (24%)	5 / 29 (17%)	$\chi^2 = 0.42$ $p = 0.52$

feedback—that is, after incorrectly classifying an object—was significant, with $\chi^2 = 14.0$, at $p < 0.001$.

5.2.2 Student-level analysis

We first present descriptive statistics of student-level NAUF in both tasks, and then examine changes in this behavior. We take three different approaches for examining such changes: (1) Comparing the NAUF rates between two halves of each of the tasks; (2) Examining the overall trend of NAUF along each task; and (3) Comparing the overall NAUF rates between the two tasks.

The average value of *not acting upon feedback* at the student-level was 0.12 in the polygons task, and 0.09 in the traffic applet. As student-level NAUF values denote the average NAUF instances across each students' actions, these values represent a relatively low rate of NAUF. However, in the polygons task, there were 17 students (59%) who demonstrated at least one NAUF instance, and in the traffic task, this number was 18 (62%).

For the polygons task, *NAUF first half* and *NAUF second half* took averages of 0.11 and 0.14, respectively; this difference is not significant. For the traffic task, *NAUF first half* and *NAUF second half* took averages of 0.13 and 0.06, respectively; this decrease is significant, with a large effect size—measured by Rank-Biserial Correlation—of 0.60. When comparing, in a between-tasks manner, first and second halves' value of NAUF, we found no significant differences.

In the polygons task, *NAUF trend* values took an average of 0.14, and in the traffic task—it took an average of 0.08. This difference is not significant. These numbers represent a slight increase in NAUF instances along the task solving process.

The difference in overall NAUF rate is not significant when compared between the tasks.

5.3 NAUF and Success (RQ3)

We tested for correlations between each of the two NAUF variables (*NAUF*, *NAUF trend*) and each of the success variables (*overall success*, *final correct classifications*). Results are listed in Table 5, and are described next.

Not acting upon feedback was significantly negatively correlated with *overall success* in both tasks; that is, the more students did not act upon feedback, the lower was their *overall success.*

Correlating with *final correct classifications*, we found a significant, negative association in the Polygons applet, and a non-significant correlation in the Traffic applet.

In the Polygons task, *NAUF trend* was negatively correlated with both *final correct classifications* and *overall success*, which means that as NAUF instances increased along the task solving process – success decreased. In the Traffic task, *NAUF trend* was significantly correlated with *overall success*, but not with *final correct classifications*.

6 Discussion

In this research, we studied how students' feedback-related behavior, while working on a digital classification task in the presence of immediate feedback, were associated with success in the task. We did so in the context of reflective symmetry, which is part of elementary school mathematics education. Our analysis was based on the learning environment log files, which allowed us to trace every student action, and therefore to portray the solution process in high resolution. In this section, we discuss our findings.

6.1 Ending tasks with no full success (RQ1)

While working on a binary classification task, immediate feedback of correctness allows students to easily derive the correct answer. Therefore, noticing feedback and acting upon it would have resulted in an overall correct classification. As we demonstrated, for some participants—this was not the case. While we cannot know whether our participants failed to notice the feedback, did not fully understand it, or intentionally decided to not act upon it (or any combination of those), this finding is aligned with the literature stating that the mere existence of feedback is not sufficient (Hattie & Timperley, 2007; Wisniewski et al., 2020). It is possible

Table 5 Summary of the associations between NAUF and success

Variable	Polygons		Traffic	
	Overall Success	Final Correct Classifications	Overall Success	Final Correct Classifications
NAUF	$\rho = -0.74$***	$\rho = -0.59$***	$\rho = -0.46$*	$\rho = -0.29$, $p = 0.13$
NAUF Trend	$\rho = 0.73$***	$\rho = -0.50$**	$\rho = -0.40$*	$\rho = -0.05$, $p = 0.80$

*p < 0.05, ***p < 0.01, ***p < 0.001

that learners simply were not able to correctly identify lines of symmetry in given objects, as was reported by other researchers (Seah & Horne, 2019); if this was the case, it is possible that our feedback did not provide enough scaffolding as for where to look for these lines. Also, our feedback did not help students to better strategize in the second task based on their behavior in the first task; indeed, this is our goal for future research; that is, identifying various relevant student behaviors—of which NAUF is one example—and designing effective feedback.

6.2 Understanding NAUF behavior (RQ2)

Our findings showed that the behavior of not acting upon feedback is existing—albeit not to a large extent—and is rather consistent and hardly changes (see 5.2.2). Therefore, we recommend that students should be explicitly taught how to make effective use of feedback. They should become familiar with the very existence of the feedback mechanism and the most effective ways of engaging with it (notice), realize what they can learn from it (understand), and recognize potential responses that would make their learning process effective (act upon). This leads us to an important recommendation in the context of learning with digital learning environments: Notice, understand, and act upon immediate feedback. Educators and designers of digital learning environments should not neglect the importance of the first two steps in this recommended process; if feedback is neither noticed nor understood it cannot be acted upon in any (meaningful) way.

It should be noted that this model of notice, understand, and act upon may work not only as deep, but also as surface behavior. However, based on the design of the task we studied here, we may cautiously deduce that students did not process the feedback as surface behavior. Mostly, this is relevant to the way reflective symmetry is understood by students, which has to do with folding a physical object around a line. As folding is not available in the digital environment, students can only implement mental folding, that is, imagining the whole process of constructing a potential line of symmetry, superimposing the two parts created by it, and testing for the fit of these two parts. This makes the task difficult, and may lead students to be convinced in their decision about an object despite feedback indicating on an erroneous classification; in such cases of discrepancy, demonstrating NAUF is indicative of deep behavior. Indeed, we found evidence that most instances of students not acting upon feedback occurred after incorrect classifications. Alternatively, we suggest two other explanations for this finding.

First, the way that the two feedback scores were positioned on the screen on top of each other may have made the incorrect score (bottom) less visible than expected. An eye-tracking study of using feedback in a digital game had shown that students spent more time on critical (negative) feedback than on confirmatory (positive) feedback (Cutumisu et al., 2019), which would mean that it is more likely for them to respond to critical than to confirmatory feedback, however apparently they did not. Relying on previous eye-tracking studies showing that the lower an information is presented on the screen, the less attention it gets (Bojar et al., 2016; Djamasbi et al., 2011). Therefore, it is possible that the higher count of not responding to feedback after correct classifications is because of the order of feedback scores and not the mere message depicted in these numbers.

Second, it may be that after wrong classifications, students are less susceptible to feedback. Students who do not perform well academically tend to attribute their failure to external factors rather than to internal ones, therefore are less susceptible to feedback that may help them improve (Coutinho et al., 2020). Also, feedback that indicates on incorrectness is likely to harm affective-motivational state (Kuklick & Lindner, 2023), and emotions have key role in feedback acceptance (Molloy et al., 2019; Naismith & Lajoie, 2018). It is therefore recommended to design feedback that sensitively elaborates why a classification is incorrect.

As for changes in not acting upon feedback, we found some evidence indicating on decrease in this behavior. This is demonstrated by a drop in the step-level measured behavior after incorrectly classifying an object, as well as a decrease in students' rate of not acting upon feedback between first and second halves of their task solving process in the Traffic task. At the same time the rate of not acting upon feedback seems to be rather consistent, as indicated by the overall lower values of the trend variable. This means that learning to use feedback effectively may take time and may require some scaffolding; indeed, using feedback has been suggested as an essential skill for today's learners (e.g., Kivunja, 2015).

6.3 NAUF and success (RQ3)

We found that not acting upon feedback is associated with lower performance, hence may be harmful, probably due to limiting opportunities to learn. An increasing trend of this behavior is also associated with lower performance. Not acting upon feedback may be seen as a form of disengagement, whether it indicates not noticing the feedback, not understanding it, or merely not taking action following it. Indeed, disengagement has been traditionally associated with low performance (e.g., Forbes-Riley & Litman, 2013; Wise & Kingsbury, 2022), and specifically in digital learning environments, e.g., in the case of gaming-the-system (Baker, 2007), or wheel-spinning (Beck & Gong, 2013).

With that, we hypothesize that the other side of that coin will also hold, that is, that acting upon feedback will be positively associated with success. It is important to emphasize that acting upon feedback is not merely a counter-behavior to the one measured here, as having no evidence for not acting upon feedback is not necessarily evidence for acting upon it.

6.4 Operationalization of feedback use

Using learning analytics yields great potential for studying student behavior in digital learning environments, as data can be collected automatically and unobtrusively via log files. This is in line with the idea of Chatti et al. (2014), which translates the main goal of learning analytics into insights, decisions and actions to improve learning and teaching. However, the question is how such data can be used to benefit mathematics learning. In doing so, we want to focus not only on the results of solving tasks, but also include data on the student's solving process. Thus, as a long-term goal, we want to investigate the effectiveness of students' solution processes in a digital environment—based on the analysis of log files. More specifically, we wish to discuss the extent to which we can empirically identify strategies and factors that lead to success.

As in any learning analytics-based study, one of the basic methodological components of our study is the operationalization of our research variables, specifically not acting upon feedback. Recalling that student-feedback interaction is a complex construct, we had to make some assumptions while trying to identify instances of not acting upon feedback in the log files. The first of these was that it would be more feasible to identify acting upon feedback than noticing it or understanding it; with only log files traces in hand (and not, for example, data from eye-tracking camera), this seems to be a reasonable assumption. Another assumption was that it would be easier to identify cases of not acting upon feedback rather than of acting upon it; following this assumption, we defined those cases where "a reasonable student" would have immediately acted, and defined a non-action when this did not happen.

6.5 Limitations

Our study is, of course, not without limitation. First, it is limited by its relatively small number of participants; we plan to increase these numbers in our future research. Second, the change in the level of difficulty between the two applets may have served as an obstacle while working through the traffic task; here we plan to design future tasks using a more moderate increase in complexity. Third, the design of our tasks may have affected student behavior—some objects might have been more prominent than others due to their location on the screen or size; in the future design of similar tasks, we will take these aspects into consideration to reduce such biases as much as possible. Finally, while relaying on log files has a major advantage both methodologically (i.e., the ability to computationally analyze data from large populations) and practically (i.e., the potential to support students with real-time feedback), our operationalization of variables should be considered only as a proxy to behaviors we wished to capture (e.g., noticing feedback); therefore, we suggest complementing such studies with another means of data collection, e.g., observations, interviews, or eye-tracking.

6.6 Conclusions

In this research, we studied student actions while working on digital binary classification tasks in the presence of immediate feedback, in the context of reflective symmetry. Overall, our findings support the notion that feedback contribution to learners is not a result of its mere existence, but rather of the way learners interact with it; therefore, our recommendations emphasize the need to educate students to work with feedback: feedback should be noticed, understood, and acted upon. This is especially relevant when learners struggle with the topic at hand, as was the case for reflective symmetry. Our recommendations though, may be applicable to any such classification tasks, across a variation of mathematical domains.

An important contribution of our research is its unique methodology following an approach from Learning Analytics by analyzing log files. An advantage of this computational approach is its ability to be implemented within digital classification tasks, hence, to identify student behavior in real-time. This way, it is possible to design and employ a real-time feedback mechanism for problem-solving strategies, holding the potential to support the learning process.

It should be noted that this approach holds the potential to be carried out on a large-scale and detect patterns from data collected in a non-obtrusive manner, that resembles authentic solution processes. The challenge in this approach lies in identifying and operationalizing variables that allow inferences on learners' understanding of a given context, which we made a first step towards. Due to the unused potential of log file analysis, we want to encourage this approach to be further explored and pursued in mathematics education.

Acknowledgements We thank the Bavarian Research Alliance and their Funding Programme for the Initiation of International Projects (BayIntAn) as well as their Scholarship Program for Bavarian and Israeli Early Career Researchers for supporting this project.

Funding Open access funding provided by Tel Aviv University. This research was supported by Bayerische Forschungsallianz.

Data availability The datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request.

Declarations

Conflict of interest The authors declare no conflict of interests.

References

Adjei, S., Ostrow, K., Erickson, E., & Heffernan, N. (2017). Clustering students in assistments: Exploring system- and school-level traits to advance personalization. *Proceedings of the 10th International Conference on Educational Data Mining, EDM 2017*, 340–341.

Baker, R. S., Corbett, A. T., Koedinger, K. R., & Wagner, A. Z. (2004). Off-task behavior in the cognitive tutor classroom: When students "game the system." *Proceedings of the 2004 Conference on Human Factors in Computing Systems - CHI '04*, 383–390.

Baker, R. S. J. D. (2007). Modeling and understanding students' off-task behavior in intelligent tutoring systems. *Conference on Human Factors in Computing Systems - Proceedings*, 1059–1068. https://doi.org/10.1145/1240624.1240785

Baker, R. S. J. D., Hershkovitz, A., Rossi, L. M., Goldstein, A. B., & Gowda, S. M. (2013). Predicting robust learning with the visual form of the moment-by-moment learning curve. *Journal of the Learning Sciences, 22*(4), 639–666. https://doi.org/10.1080/10508406.2013.836653

Baker, R. S., & Koedinger, K. R. (2018). Towards demonstrating the value of learning analytics. In D. Niemi, R. D. Pea, B. Saxberg, & R. E. Clark (Eds.), *Learning analytics in education* (pp. 49–62). Information Age Publishing.

Beck, J. E., & Gong, Y. (2013). *Wheel-spinning: Students who fail to master a skill* (pp. 431–440). https://doi.org/10.1007/978-3-642-39112-5_44

Ben-Haim, E., Cohen, A., & Tabach, M. (2019). Types of graphic interface design and their role in learning via mathematical applets at the elementary school. *Eleventh Congress of the European Society for Research in Mathematics Education*.

Bojar, O., Děchtěrenko, F., & Zelenina, M. (2016). A pilot eye-tracking study of WMT-style ranking evaluation. In G. Rehm, A. Burchardt, O. Bojar, C. Dugast, M. Federico, J. van Genabith, B. Haddow, J. Hajič, K. Harris, P. Koehn, M. Negri, M. Popel, L. Specia, M. Turchi, & H. Uszkoreit (Eds.), *Proceedings of the LREC 2016 Workshop "Translation Evaluation: From Fragmented Tools and Data Sets to an Integrated Ecosystem"* (pp. 20–26).

Bruder, R., Brunner, E., & Siller, H. S. (2021). Unterrichtsforschung unter fachlichen Perspektiven – Mathematik. *Handbuch Schulforschung*. https://doi.org/10.1007/978-3-658-24734-8_49-1

Cavalcanti, A. P., Barbosa, A., Carvalho, R., Freitas, F., Tsai, Y. S., Gašević, D., & Mello, R. F. (2021). Automatic feedback in online learning environments: A systematic literature review. *Computers and Education: Artificial Intelligence, 2*, 100027. https://doi.org/10.1016/J.CAEAI.2021.100027

Chatti, M. A., Lukarov, V., Thüs, H., Muslim, A., Yousef, A. M. F., Wahid, U., Greven, C., Chakrabarti, A., & Schroeder, U. (2014). Learning analytics: Challenges and future research directions. *Eleed*, *10*(1). https://www.eleed.de/archive/10/4035/index_html

Colthorpe, K., Zimbardi, K., Ainscough, L., & Anderson, S. (2015). Know thy student! Combining learning analytics and critical reflections to increase understanding of students' self-regulated learning in an authentic setting. *Journal of Learning Analytics, 2*(1), 134–155.

Coutinho, M. V. C., Thomas, J., Fredricks-Lowman, I., & Verlinden Bondaruk, M. (2020). The Dunning-Kruger effect in Emirati college students: Evidence for generalizability across cultures. *International Journal of Psychology and Psychological Therapy, 20*(1), 29–36.

Cutumisu, M., Turgeon, K.-L., Saiyera, T., Chuong, S., González Esparza, L. M., MacDonald, R., & Kokhan, V. (2019). Eye tracking the feedback assigned to undergraduate students in a digital assessment game. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2019.01931

de Kock, W. D., & Harskamp, E. G. (2016). Procedural versus content-related hints for word problem solving: An exploratory study. *Journal of Computer Assisted Learning, 32*(5), 481–493. https://doi.org/10.1111/jcal.12148

Djamasbi, S., Sigel, M., & Tullis, T. (2011). Visual hierarchy and viewing behavior: An eye tracking study. *Human-Computer Interaction, Design and Development Approaches, Lecture Notes in Computer Science*, 331–340.

Drullion, F. (2019). Immediate feedback assessment technique in a mathematics course for non technical majors. *SoTL Commons Conference*. https://digitalcommons.georgiasouthern.edu/sotlcommons/SoTL/2019/70

Duval, E., & Verbert, K. (2012). Learning analytics. *Eleed, 8*(1), 96–99. https://doi.org/10.1109/MC.2007.331

Fancsali, S. E., & Ritter, S. (2014). Context personalization, preferences, and performance in an intelligent tutoring system for middle school mathematics. *ACM International Conference Proceeding Series, Figure, 1*, 73–77. https://doi.org/10.1145/2567574.2567615

Ferguson, R. (2012b). *The state of learning analytics in 2012: A review and future challenges (technical report KMI-12–01)*.

Ferguson, R. (2012a). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning, 4*(5/6), 304–317. https://doi.org/10.1504/IJTEL.2012.051816

Forbes-Riley, K., & Litman, D. (2013). When does disengagement correlate with performance in spoken dialog computer tutoring? *International Journal of Artificial Intelligence in Education, 22*(1–2), 39–58.

Frey, C. (2022). Flipped classroom in Introductory Mathematics: Data analysis of student success using immediate feedback vs. deferred feedback. *SoTL Commons Conference*. https://digitalcommons.georgiasouthern.edu/sotlcommons/SoTL/2022/102

Gal, T., & Hershkovitz, A. (2019). Different types of response-based feedback in mathematics: The case of textual and symbolic messages. *ACM International Conference Proceeding Series*. https://doi.org/10.1145/3303772.3303815

Golke, S., Dörfler, T., & Artelt, C. (2015). The impact of elaborated feedback on text comprehension within a computer-based assessment. *Learning and Instruction, 39*, 123–136. https://doi.org/10.1016/J.LEARNINSTRUC.2015.05.009

Götz, D., & Gasteiger, H. (2022). Reflecting geometrical shapes: Approaches of primary students to reflection tasks and relations to typical error patterns. *Educational Studies in Mathematics*. https://doi.org/10.1007/s10649-022-10145-5

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. https://doi.org/10.3102/003465430298487

Hershkovitz, A., Tzayada, O., Ezra, O., Cohen, A., Tabach, M., Levy, B., Segal, A., & Gal, K. (2019). Can an algorithm prepare students for tasks without knowing what the tasks are? *Proceedings - 6th Annual Conference on Computational Science and Computational Intelligence, CSCI 2019*, 754–759. https://doi.org/10.1109/CSCI49370.2019.00143

Hershkovitz, A., Tabach, M., Noster, N., & Siller, H.-S. (2023). Student behavior while engaged with feedback-enhanced digital sorting tasks. In M. Ayalon, B. Koichu, R. Leikin, L. Rubel, & M. Tabach (Eds.), *Proceedings of the 46th Conference of the International Group for the Psychology of Mathematics Education (Volume 3)* (pp. 51–28). PME.

Hershkovitz, A., de Baker, R. S. J., Gobert, J., Wixon, M., & Pedro, M. S. (2013). Discovery with models: A case study on carelessness in computer-based science inquiry. *American Behavioral Scientist, 57*(10), 1480–1499. https://doi.org/10.1177/0002764213479365

Hershkovitz, A., & Nachmias, R. (2011). Online persistence in higher education web-supported courses. *Internet and Higher Education, 14*(2), 98–106. https://doi.org/10.1016/j.iheduc.2010.08.001

Hershkovitz, A., Tabach, M., & Cohen, A. (2021). Online activity and achievements in elementary school mathematics: A large-scale exploration. *Journal of Educational Computing Research*. https://doi.org/10.1177/07356331211027822

Iraj, H., Fudge, A., Khan, H., Faulkner, M., & Pardo, A. (2021). Narrowing the feedback gap: Examining student engagement with personalized and actionable feedback messages. *Journal of Learning Analytics, 8*(3), 101–116. https://doi.org/10.18608/jla.2021.7184

Kivunja, C. (2015). Teaching students to learn and to work well with 21st century skills: Unpacking the career and life skills domain of the new learning paradigm. *International Journal of Higher Education, 4*(1), 1–11.

Kolbinger, K.-H., Kriegelstein, W., Ernst, G., Kistella, A., Leininger, P., & Wallrabenstein, H. (2006). *Nussknacker – Unser Rechenbuch. 3. Schuljahr [in German]*. Klett.

Kuikka, M., Laakso, M.-J., & Joshi, M. (2016). The effect of the immediate feedback by the collaborative education tool ViLLE on learning for business mathematics in higher education. *Journal of Educational Technology Systems, 45*(1), 34–49. https://doi.org/10.1177/0047239515625887

Kuklick, L., & Lindner, M. A. (2023). Affective-motivational effects of performance feedback in computer-based assessment: Does error message complexity matter? *Contemporary Educational Psychology, 73*, 102146. https://doi.org/10.1016/j.cedpsych.2022.102146

Leiba, M. (2010). Assessing mathematical problem solving behavior in web-based environments using data mining. *CEUR Workshop Proceedings, 709*, 37–42.

Lomas, D., Patel, K., Forlizzi, J. L., & Koedinger, K. R. (2013). Optimizing challenge in an educational game using large-scale design experiments. *Conference on Human Factors in Computing Systems - Proceedings*, 89–98. https://doi.org/10.1145/2470654.2470668

McGuire, P., Tu, S., Logue, M. E., Mason, C. A., & Ostrow, K. (2017). Counterintuitive effects of online feedback in middle school math: Results from a randomized controlled trial in ASSISTments. *Educational Media International, 54*(3), 231–244. https://doi.org/10.1080/09523987.2017.1384161

Molloy, E., Noble, C., & Ajjawi, R. (2019). Attending to emotion in feedback. In M. Henderson, R. Ajjawi, D. Boud, & E. Molloy (Eds.), *The impact of feedback in higher education improving assessment outcomes for learners* (pp. 83–106). Palgrave Macmillan.

Naismith, L. M., & Lajoie, S. P. (2018). Motivation and emotion predict medical students' attention to computer-based feedback. *Advances in Health Sciences Education, 23*(3), 465–485. https://doi.org/10.1007/s10459-017-9806-x

Narciss, S. (2013). Designing and evaluating tutoring feedback strategies for digital learning environments on the basis of the Interactive Tutoring Feedback Model. *Digital Education Review, 23*, 7–26.

Nitsch, Renate. (2015). *Diagnose von Lernschwierigkeiten im Bereich funktionaler Zusammenhänge : eine Studie zu typischen Fehlermustern bei Darstellungswechseln*.

Noster, N., Hershkovitz, A., Siller, Hans.-S., & Tabach, M. (2022a). Students' strategies for identifying reflective symmetry of extra-mathematical shapes in a digital environment. *ERME Topic Conference on Mathematics Education in the Digital Age*.

Noster, N., Hershkovitz, A., Tabach, M., & Siller, H.-S. (2022b). Learners' strategies in interactive sorting tasks. In I. Hilliger, P. J. Muñoz-Merino, T. De Laet, & A. F. T. Ortega-Arranz (Eds.), *Educating for a New Future: Making Sense of Technology-Enhanced Learning Adoption. EC-TEL 2022. Lecture Notes in Computer Science, vol 13450* (pp. 285–298). Springer.

Panadero, E., & Lipnevich, A. A. (2022). A review of feedback models and typologies: Towards an integrative model of feedback elements. *Educational Research Review, 35*, 100416. https://doi.org/10.1016/J.EDUREV.2021.100416

Papamitsiou, Z., & Economides, A. A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. In *Educational Technology & Society* (Vol. 17, Issue 4).

Polya, G. (1981). *Mathematical discovery: On understanding learning and teaching problem solving*. John Wiley & Sons.

Roder, U. (2020). *Ein Förderkonzept zu mathematischem Grundwissen und Grundkönnen am Übergang in die Sekundarstufe II.*https://doi.org/10.1007/978-3-658-28118-2

Rosenthal, I., Ginat, H., & Gafni, R. (2010). *Simply mathematics: Mathematics for elementary school (Book 2) [in Hebrew]* (R. Steinberg, Ed.). Kinneret Publishing House .

Schimmel, B. J., & Schimmel, B. J. (1988). Patterns in students' selection of feedback in computer-based instruction. *Presented at the Annual Meeting of the American Educational Research Association*.

Seah, R., & Horne, M. (2019). An exploratory study on students' reasoning about symmetry. *Mathematics Education Research: Impacting Practice: Proceedings of the 42nd Annual Conference of the Mathematics Education Research Group of Australasia*, 628–635.

Timmers, C., & Veldkamp, B. (2011). Attention paid to feedback provided by a computer-based assessment for learning on information literacy. *Computers & Education, 56*(3), 923–930. https://doi.org/10.1016/J.COMPEDU.2010.11.007

Tsamir, P., Tirosh, D., & Levenson, E. (2008). Intuitive nonexamples: The case of triangles. *Educational Studies in Mathematics, 69*(2), 81–95. https://doi.org/10.1007/s10649-008-9133-5

Velasquez, N. F., Martin, T., & Maughan, J. (2014). Learning aid use patterns and their impact on exam performance in online developmental mathematics. *Proceedings of the 7th International Conference on Educational Data Mining (EDM)*, 379–380.

Winstone, N. E., Nash, R. A., Parker, M., & Rowntree, J. (2016). Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of recipience processes.

Educational Psychologist, 52(1), 17–37. https://doi.org/10.1080/00461520.2016.1207538

Wise, S. L., & Kingsbury, G. G. (2022). Performance decline as an indicator of generalized test-taking disengagement. *Applied Measurement in Education, 35*(4), 272–286. https://doi.org/10.1080/08957347.2022.2155651

Wisniewski, B., Zierer, K., & Hattie, J. (2020). The Power of feedback revisited: A meta-analysis of educational feedback research. In *Frontiers in Psychology* (Vol. 10, p. 3087). Frontiers Media S.A. https://doi.org/10.3389/fpsyg.2019.03087

Wixon, M., Baker, R. S. J. d., Gobert, J. D., Ocumpaugh, J., & Bachmann, M. (2012). *WTF? Detecting Students Who Are Conducting Inquiry Without Thinking Fastidiously* (pp. 286–296). https://doi.org/10.1007/978-3-642-31454-4_24

Xie, J., Mojarad, S., Shubeck, K., Essa, A., Baker, R. S., & Hu, X. (2017). Student learning strategies and behaviors to predict success in an online adaptive mathematics tutoring system. *Proceedings of the 10th International Conference on Educational Data Mining, EDM 2017*, 460–465.