



<http://www.diva-portal.org>

Preprint

This is the submitted version of a paper presented at *Conference of the International Group for the Psychology of Mathematics Education*.

Citation for the original published paper:

Schindler, M., Schaffernicht, E., Lilienthal, A J. (2020)

Identifying student strategies through eye tracking and unsupervised learning: The case of quantity recognition

In: Inprasitha, M., Changsri, N. & Boonsena, N. (ed.), *Interim Proceedings of the 44th Conference of the International Group for the Psychology of Mathematics Education. Khon Kaen, Thailand: PME*. (pp. 518-527).

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:oru:diva-89112>

IDENTIFYING STUDENT STRATEGIES THROUGH EYE TRACKING AND UNSUPERVISED LEARNING: THE CASE OF QUANTITY RECOGNITION

Maike Schindler¹, Erik Schaffernicht², Achim J. Lilienthal²

¹University of Cologne

²Örebro University

Identifying student strategies is an important endeavor in mathematics education research. Eye tracking (ET) has proven to be valuable for this purpose: E.g., analysis of ET videos allows for identification of student strategies, particularly in quantity recognition activities. Yet, “manual”, qualitative analysis of student strategies from ET videos is laborious—which calls for more efficient methods of analysis. Our methodological paper investigates opportunities and challenges of using unsupervised machine learning (USL) in combination with ET data: Based on empirical ET data of $N = 164$ students and heat maps of their gaze distributions on the task, we used a clustering algorithm to identify student strategies from ET data and investigate whether the clusters are consistent regarding student strategies.

INTRODUCTION

For researchers and practitioners (e.g., teachers) in mathematics education, it is important to not only evaluate student achievements, their results and products, but also to analyze students' thought processes and individual strategies leading to such products. In recent years, eye tracking (ET)—the recording of eye movements—has gained increasing importance in mathematics education research (Lilienthal & Schindler, 2019). Among others, it has proven to be valuable to analyze student strategies in different mathematical areas (e.g., Bruckmeier et al., 2019; Obersteiner & Tumpek, 2016), including quantity recognition in whole number representations (Lindmeier & Heinze, 2016; Schindler & Lilienthal, 2018). For example, Schindler et al. (2019a) analyzed student strategies in determining quantities in the 100-dot field and 100-abacus based on ET data: They used gaze-overlaid videos (videos of the scene with the eye gaze visualized as dot wandering around) to infer student strategies. However, such qualitative analysis of ET data is laborious: Analyzing ET data, which are rich by nature, is time-consuming and demanding (Klein & Ettinger, 2019). This calls for more efficient methods of analysis when bigger numbers of students are studied, and student strategies are to be inferred (Klein & Ettinger, 2019).

Our methodological paper explores the possibility to identify student strategies in whole number representations using ET data combined with unsupervised

machine learning (USL). Based on data from $N = 164$ fifth grade students, we use a clustering algorithm (a specific instance of USL), to investigate the possibility to identify student quantity recognition strategies from so-called gaze heat maps (see Fig. 2). Broadly, we investigate what opportunities and challenges USL offers for identifying quantity recognition strategies. In particular, we ask the question: *Does the USL provide consistent clusters with respect to student strategies?*

Our paper illustrates with examples how a clustering algorithm, applied to heat maps, can be used to identify student strategies (“proof of concept”). We investigate the consistency of the clusters provided by the USL through qualitative interpretation using qualitative previous findings and elaborate on opportunities and challenges of USL.

EYE TRACKING IN MATHEMATICS EDUCATION RESEARCH

Eye tracking allows for a recording of spatio-temporal sequences of gaze points that indicate visual attention. The connection between gaze and visual attention exists due to an economic feature of the human eye, which concentrates a substantial fraction of the receptors on the retina in the small area of the *fovea*. Thus, in order to pay attention in detail, humans need to move their eyes constantly so that the area of interest is in line with the fovea, a process that can be tracked with ET devices unobtrusively by visually observing the pupils. ET is of interest for mathematics education research since the recorded sequences of gaze points do allow inferences about mental processes, though interpretation of gaze movements is not straightforward and bijective (Schindler & Lilienthal, 2019). ET is of growing interest since ET devices became increasingly affordable, advanced, and accurate (Lilienthal & Schindler, 2019); due to theoretical advances in interpretation (Schindler & Lilienthal, 2019); and since the required computational resources for partially automated analysis are available at low cost, which makes ET applications using (partially) automated analysis available for research and, in the future, also for mathematics education practitioners (e.g., teachers).

MACHINE LEARNING

The term *Machine Learning* (ML) refers to a set of methods for automated analysis of data, specifically “methods that can automatically detect patterns in data, and then use the uncovered patterns to predict future data, or to perform other kinds of decision making under uncertainty” (Murphy, 2012, p. 1). There are two major types of ML: *Supervised learning* (SL) algorithms learn a mapping between training samples and respective output. This means that each sample in the training set must be labelled. The learned mapping can then be used to make categorical or nominal predictions (Murphy, 2012). SL is thus also called *predictive learning*. SL is used, for example, in Schindler et al.’s (2019b) study, where the training samples are (as in this paper) ET sequences

represented in the form of heat maps, with labels that specify each heat map to belong to a student with or without mathematical difficulties. After training, the SL algorithm can be used to classify previously unseen heat maps and predict whether the corresponding student has mathematical difficulties or not.

The second major type of ML is *unsupervised learning (USL)* where only training samples but no labels are given. The computer is then tasked to “find ‘interesting patterns’ in the data” (Murphy, 2012, p. 2). This is also called *knowledge discovery*. As Murphy (2012) notes, USL is a much less well-defined problem than SL. In this paper, we use *clustering*, a form of USL in which the set of samples (here: gaze heat maps) is divided (“clustered”) into a number of groups. A clustering algorithm tries to find a meaningful division of the input data, but how a good division may look like and the “correct” number of clusters is not known *a priori*. To the best of our knowledge, USL has not been used on ET data in mathematics education research so far.

QUANTITY RECOGNITION IN WHOLE NUMBER REPRESENTATIONS

Whole number representations such as the 100 dot field or the 100 abacus (also called “100-frame”), which visualize substructures of 100 (50, 10s, 5s), are often used for students to learn the number range up to 100 (Gaidoschik, 2015). Previous research has shown that students, when perceiving quantities in such representations, make use of structures such as 10s (rows) and 5s (Obersteiner et al., 2014). While the analysis of student strategies in such representations is challenging (Obersteiner et al., 2014), recent studies have indicated that ET is a useful tool to identify strategies, e.g., from ET videos (Schindler & Lilienthal, 2018) or scan-paths, which indicate where the students looked at (Lindmeier & Heinze, 2016). Whereas such studies using ET to identify strategies are promising, the qualitative analysis of gaze patterns is demanding and time-consuming—especially for empirical studies with larger numbers of participants. Therefore, we investigate the opportunities that USL may offer to help identify student strategies based on their spatial gaze distributions on the task.

THIS STUDY

Participants. We use data from a study with 164 fifth-grade students (92 boys, 72 girls) in a German comprehensive school. The mean age was 10;9 (SD = 0;7). The study took place in the first weeks of fifth grade. Using a standardized arithmetic paper-pencil test, we identified 59 children as typically developing in mathematics, 69 children to encounter mathematical difficulties, and 36 to be “at risk” to have mathematical difficulties (see Schindler et al., 2019a;b for a detailed description of the test).

Tasks, procedure, and eye tracker. We used a digital version of the 100-dot field. We used the same numbers as in Schindler et al. (2019a), where student

strategies were inferred from ET videos qualitatively (7, 15, 20, 31, 43, 54, 68, 76, 89, 92, and 100; in randomized order). The students were tested individually. We used Tobii x3-120, a remote eye tracker at a sampling rate of 120 Hz, which was mounted at the bottom edge of the 24" full HD computer monitor. It was calibrated through a nine-point calibration. Before the students worked on the tasks, they saw a picture of the dot field and were to describe it. The students got two practice tasks (with numbers not used in further tasks). They were instructed to always name the number of dots as fast and correctly as possible. Before each task, the students were asked to fixate a star in the middle of the screen. The students did not receive a response on the correctness of their answers. We made audio recordings of verbal answers.

Heat maps. ET provides rich information and a large amount of data, reflecting that gaze patterns can differ in multiple ways. To find groups of strategies ("clusters"), we chose a representation of the recorded gazes to facilitate the subsequent analysis. This representation needed to reduce the amount of data the clustering algorithm has to handle while preserving the relevant features of the gaze patterns. Based on previous research that indicated a variety of student gaze distributions on the task sheets in quantity recognition tasks (Schindler et al., 2019a), we decided to use heat maps that show the spatial distribution of gazes over the presented digital task sheets integrated over the whole duration of a task. We used the Tobii Pro Lab Software to produce individual student heat maps. For clustering, we included only heat maps of correctly or inversely solved (common mistake in German, e.g., for 89: "ninety-eight") tasks to assure that the students actually perceived the given information rather than guessed. In case of 89 on the dot field (focus of the *Results Section*), 90 heat maps were included.

Clustering. To automatically determine groups of similar heat maps, a *definition for the (dis-)similarity* between two heat maps is required. We use the Euclidean distance between the images: The sum of the squared pixel differences between two heat maps measures dissimilarity (Goshtasby, 2012). Calculating the Euclidean distance is a standard approach to determine similarity between images in digital image processing.

A second important choice concerns the *clustering algorithm* that assigns groups based on the similarity of heat maps. We use self-organizing maps (SOMs) (Kohonen, 2001), which are suited for explorative data analysis (Kaski, 1997). SOMs do not automatically determine the number of groups present in the data (which is a very hard problem) but require the number as input parameter. Since previous empirical work hinted at a set of five different kinds of strategies for quantity recognition in whole number representations (Schindler et al., 2019a), we use a structure with nine clusters, arranged in a 3x3 grid. Using nine clusters allows for the possibility that the algorithm would identify more strategies than previously found—or to differentiate them further.

SOMs have the rather unusual feature that they assume an *a priori* topology over the relationship between the different groups. While this does not necessarily guarantee for optimal clustering results, the topology, usually a 2D grid (Fig. 1), provides an additional tool to interpret the clustering results: neighborhood indicates similarity. This study utilizes the SOM algorithm implemented in the Matlab Deep Learning Toolbox with a hexagonally connected 3x3 grid and default parameters. In the clustering process, each of the student heat maps is assigned to one of the nine clusters. These assignments are iteratively optimized until all similar heat maps are assigned to the same cluster, while highly dissimilar heat maps are assigned to different clusters on opposite ends of the 3x3 grid. As a result, each heat map is assigned to a group that contains its most similar peers. The implicit assumption here is that due to the similarity of the heatmaps in each group these groups represent particular quantity recognition strategies. For each cluster, we calculate a cluster prototype as the average of all heat maps assigned to that cluster (Fig. 1). These average heat maps help to draw conclusions about the quantity recognition strategy that every cluster may represent.

Analyzing the clusters. To answer the question if USL provides consistent clusters with respect to student strategies, for every task we regard each cluster of the SOM and qualitatively assign a tentative strategy based on the average heat map. We then analyze all single heat maps in each cluster: In particular, we qualitatively assign a strategy to each heat map, based on the set of strategies found through qualitative analyses by Schindler et al. (2019a): (1) *counting all*, where the students counted all dots shown, (2) *counting fives*, where the students counted groups of fives, (3) *counting rows*, where students counted all rows displayed, (4) *using 50 as unit*, e.g., when determining 76, they perceived 50 in one glance and counted only the further rows, and (5) *subtraction/last row*, where the students, e.g., in 89 looked at the missing 90st dot, or only on the last row of displayed dots. Note that in Schindler et al.'s (2019a) study, the design was alike to ours: This applies to the (identical) tasks, the procedure, ET, etc. The participants were at the same age and also at the beginning of fifth grade. The main difference is that Schindler et al. investigated only 20 students (whereof 10 were found to have MD). Because of the larger number of 164 students in our study, we assume that our data set may include all strategies found by Schindler et al. (2019a).

RESULTS

In the following, we will pursue the question: *Does the USL provide consistent clusters with respect to student strategies?* We do so by using one task as an example: 89 on the dot field. We use this particular task, since it affords a variety of strategies (Schindler et al., 2019a) and, thus, is an interesting case for the clustering.

For the task 89, the USL found four substantial clusters (Fig. 1), whereas five clusters remained effectively empty with only one member heat map that can be considered an outlier. Regarding the average heat maps of the clusters (Fig. 1, right), we tentatively assigned strategies to these four clusters: (7) *Counting Rows on the Right*, (9) *Counting Rows in the Middle*, (1) *Last Row/Subtraction*, and (3) *Counting Rows on the Left*.

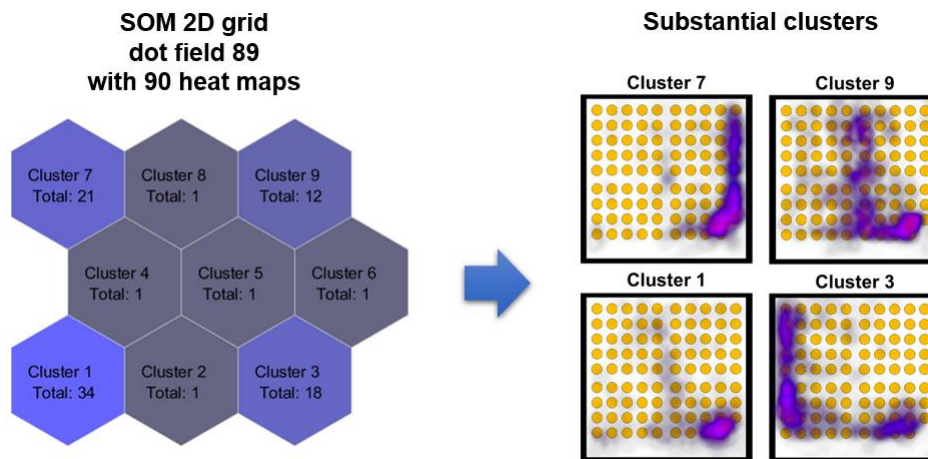


Figure 1: SOM for task 89 dot field (left) and all substantial clusters (with $n > 1$) visualized through their average heat map prototype (right).

Cluster 7: “Counting Rows on the Right” ($n=21$). Of the 21 heat maps in this cluster, we identified 19 heat maps to indicate the strategy *counting rows*, which is consistent with the impression from the average heat map: The gazes are in every row, and the pattern indicates a counting process (Fig. 2). The heat maps indicate that these 19 students counted at the right edge of the rows. The remaining two heat maps in this cluster correspond to the strategy *using 50*: Here, there are no/few gazes on the upper half of the dot field, and the gaze patterns indicate that the students counted rows 6 to 9 at the right edge of the respective rows (Fig. 2). The similarity in appearance with a concentration at the right edge of the rows in the lower half of the dot field is likely the explanation why the USL put the two *using 50*-heat maps together with the 19 heat maps that indicate *counting (all) rows*. The clustering result is reasonable since in any instance there was presumably (at least some) counting of rows on the right side.

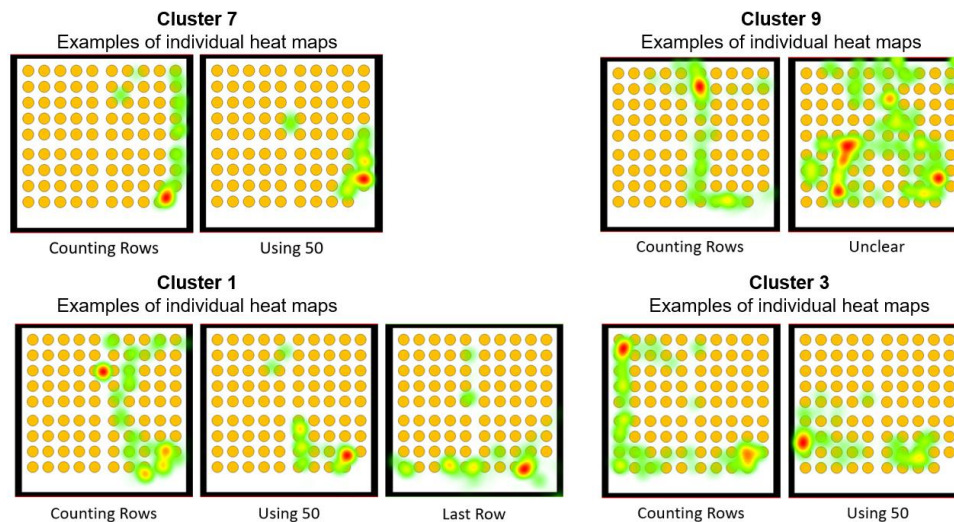


Figure 2: Examples of individual heat maps

Cluster 9: “Counting Rows in the Middle” ($n=12$). Of the 12 heat maps in this cluster, we found 7 heat maps to reflect the strategy *counting rows*, consistent with the assignment to the cluster prototype. The counting pattern is situated in the middle of the dot field, indicating that the students counted rows in the middle (Fig. 2). For the other 5 heat maps in this cluster, we are unable to identify a clear strategy. They were marked as “*unclear*” (Fig. 2): The gazes are spread over the task sheet, possibly reflecting a multitude of strategies. An indication that this cluster may contain a variety of different strategies is the rather noisy appearance of the cluster prototype.

Cluster 3: “Counting Rows on the Left” ($n=13$). 8 heat maps in this cluster indicate the strategy *counting rows*, with the gazes at the left edge of the rows (see Fig. 2). The other 5 heat maps indicate *use of 50*, since there are hardly any gazes on the upper 50 dots, but gazes that indicate that the students counted the rows from the 6th row onwards at the left edge. Similar to Cluster 7, this explains why these two kinds of heat maps were both included in the same cluster: The patterns were similar in a way that the gaze density at the left edge is high.

Cluster 1: “Last Row/Subtraction” ($n=34$). For this cluster, we found three different kinds of strategies: 7 of the heat maps indicated that the students *counted rows* (see Fig. 2). In 11 cases, we identified *using 50*: The students’ gazes indicated that the students counted rows 6 to 9 (Fig. 2). Finally, 15 heat maps indicated that the students focused only on the *last row* displayed (Fig. 2) or that they focused only on the missing 90st point, indicating a *subtraction* strategy. We assume that this relates to the distance metric used, which regards the intensity of the gaze distribution: Since the areas of the dot field that are different between these strategies have a relatively low intensity (light green), but all heat maps in this cluster have a common feature, the “blob” in the right

corner of the last row, which is intense (warm colors), this “blob” may be decisive here.

Answering the research question *if the USL provides consistent clusters with respect to student strategies*, we can say that the clusters found were—in the used example of 89 on the dot field—consistent in a certain way, but different from our previous qualitative analyses. For example, for the USL, heat maps reflecting *counting rows on the right* and *using 50* are similar and belong to one cluster if students when using 50 count the rows 6 to 9 on the right side. On the other hand, *counting rows on the right* and *counting rows on the left* belong to two different clusters. The clustering algorithm operates on visual similarity of heatmaps and inherently cannot cluster strategies together that manifest themselves very differently in the gaze distribution. A second important observation is that in cases where a student strategy involves different processes (e.g., grasping 50 in a glance and counting rows 6 to 9), clustering cannot evaluate what process is decisive for the strategy—as it was done in our previous study (Schindler et al., 2019a). Yet, given that the clusters found in our approach seldom involved more than two strategies, we find that they are—to a certain extent—consistent with respect to student strategies. So, if a student heat map belongs to one cluster, one can say that the student most likely had one or another strategy.

DISCUSSION

In this paper, we explore the possibility to identify student strategies in whole number representations using ET combined with USL. Based on ET data from $N = 164$ fifth grade students, we use the SOM algorithm for clustering and ask whether this automated analysis *provides consistent clusters with respect to student strategies*. Our question relates to a fundamental issue of USL: Compared to SL, where it is possible to quantify the performance of the trained algorithm for classification, there is no obvious error metric for USL (Murphy, 2012). As error metric from the application domain of mathematics education, we tested whether clustering identifies consistent groups regarding the strategies they represent. We found that this is true only to some extent. This is understandable: Our clustering of heat maps compares solely the visual appearance of the quantity recognition process as a whole and thus inherently cannot decompose strategies or give higher weight to certain features (e.g., the absence of gazes on the upper half). One would rather expect to find more clusters than possible strategies, since different combinations of strategies could result in additional, likely more consistent clusters. We did *not* observe such an “over-clustering” tendency and it will be subject of future work to evaluate whether other clustering algorithms and the use of other distance metrics result in a higher number and more consistent clusters.

We would like to stress that this paper gives an example of an empirical study in which Artificial Intelligence (AI) is used to support human researchers. Here, essentially, the AI component provides an independent view on a data set and makes suggestions about meaningful partitioning of the data. Human researchers interpret and verify these suggestions based on pre-studies with smaller numbers of participants and a principle understanding of the applied ML algorithms. Indeed, the clusters identified in this paper have predominantly a clear interpretation, which may be meaningful in some contexts and clearly provided an independent view from a different angle.

References

- Bruckmeier, G., Binder, K., Krauss, S., & Kufner, H.-M. (2019). An eye-tracking study of statistical reasoning with tree diagrams and 2 x 2 tables. *Frontiers in Psychology, 10*, 632.
- Gaidoschik, M. (2015). Einige Fragen zur Didaktik der Erarbeitung des „Hunderterraums“. *Journal für Mathematik-Didaktik, 36*(1), 163–190.
- Goshtasby A.A. (2012). *Image registration*. London: Springer.
- Kaski, S. (1997). *Data exploration using self-organizing maps*. Doctoral Thesis. Helsinki University of Technology: Helsinki.
- Klein, C., & Ettinger, U. (2019). *Eye movement research: An introduction to its scientific foundations and applications*. Cham, Switzerland: Springer.
- Kohonen, T. (2001). *Self-Organizing Maps* (3rd. ed.). Berlin: Springer.
- Lindmeier, A., & Heinze, A. (2016). *Strategies for recognizing quantities in structured whole number representations – A comparative eye-tracking study*. Paper presented at 13th International Congress on Mathematical Education (ICME-13), 2016.
- Lilienthal, A.J., & Schindler, M. (2019). Eye tracking research in mathematics education: A PME literature review. In *Proceedings of the 43rd Conference of the IGPME* (Vol. 4, p. 62). Pretoria, South Africa: PME.
- Murphy, K.P. (2012). *Machine learning: A probabilistic perspective*. Cambridge, MA: MIT.
- Obersteiner, A., Reiss, K., Ufer, S., Luwel, K., & Verschaffel, L. (2014). Do first graders make efficient use of external number representations? The case of the twenty-frame. *Cognition and Instruction, 32*(4), 353–373.
- Obersteiner, A., & Tumpek, C. (2016). Measuring fraction comparison strategies with eye-tracking. *ZDM, 48*, 255–266.
- Schindler, M., Bader, E., Lilienthal, A.J., Schindler, F., & Schabmann, A. (2019a). Quantity recognition in structured whole number representations of students with mathematical difficulties: An eye-tracking study. *Learning Disabilities: A Contemporary Journal, 17*, 5–28.

- Schindler, M., & Lilienthal, A.J. (2019). Domain-specific interpretation of eye tracking data: Towards a refined use of the eye-mind hypothesis for the field of geometry. *Educational Studies in Mathematics*, *101*, 123–139.
- Schindler, M., & Lilienthal, A.J. (2018). Eye-tracking for studying mathematical difficulties
—also in inclusive settings. In *Proceedings of the 42nd Conference of the IGPME* (Vol. 4, pp. 115–122). Umeå, Sweden: PME.
- Schindler, M., Schaffernicht, E., & Lilienthal, A. (2019b). Differences in quantity recognition of students with and without mathematical difficulties: Analysis through ET and AI. In *Proceedings of the 43rd Conference of the IGPME* (Vol. 3, pp. 281–288). Pretoria, South Africa: PME.