TOWARDS AUTOMATED GROUPING: UNRAVELING MATHEMATICS TEACHERS' CONSIDERATIONS

Rotem Abdu[1], Shai Olsher[2], & Michal Yerushalmy[3]

University of Haifa, Israel; rotem_abdu@yahoo.com[1], olshers@edu.haifa.ac.il[2], michalyr@edu.haifa.ac.il[3]

ABSTRACT

What are mathematics teachers' considerations in grouping students, and how could automated formative assessment systems help them in doing it? In this study, nine teachers were asked to use data on students' performance in a mathematics task, derived from an automated formative assessment system, to create pairs in which students could contribute to their peers. We called this grouping strategy "complementary." The teachers were also asked to explain their considerations for each grouping. We found two main grouping strategies in addition to the complementary one: based on similar answers ("similarity"), and based on dissimilar answers, in which one student performed better than another and could teach the other ("hierarchy"). Findings show that despite the experimenter's request to group students based on complementarity, teachers mostly grouped based on other considerations, at times even grouping students whose answers were complementary using hierarchical considerations. In some cases, different teachers formed the same groups of students based on different grouping strategies. The findings confirm the hypothesis that informed grouping may be challenging for teachers, and may benefit greatly from an automated pairing system.

Keywords: informed grouping; differentness; linear functions; personal example space; automated formative assessment.

INTRODUCTION

How do mathematics teachers group students? What kind of information should they use and how may this information help them in forming groups? What are their considerations as they do so? This information is crucial to our quest, in which we aim to use data about students' mathematical work, generated by an automated formative assessment system, to perform informed grouping of students. Thus, we asked mathematics teachers to groups students using data retrieved from an automated mathematics formative assessment system, and explain their reasons for these groupings. The paper begins with a literature review on formative assessment, collaborative learning, and informed grouping. It continues with a description of the study, followed by our main findings. We end with a discussion of the findings, with particular focus on how these results may influence the design and research of future automated grouping systems.

THEORETICAL BACKGOUND

Group learning

Group learning is a fundamental pedagogical practice (Dillenbourg, 1999), although it is not always effective (Barron, 2003; Hoyles, Healy & Sutherland, 1991) and requires teachers to take into account various considerations (Webb, 2009). An important practice is informed grouping, allocating students to groups based on predefined dimensions and knowledge of learners' abilities along these dimensions. Researchers who design and study informed grouping are concerned with the different dimensions along which grouping should be carried out, and how these dimensions may best inform grouping for it to be beneficial to students.

Groups of students can be either homogeneous or heterogeneous (e.g., Maqtary, Mohsen, & Bechkoum, 2017). A homogeneous group is usually comprised of learners who are considered similar along some measures such as age, gender, or grades. When students learn in a homogenous group, it is more likely that they make progress at the same pace and are able to solve similar tasks of about the same difficulty (e.g., Connor et al., 2013). There are two types of heterogeneous grouping: hierarchic and complementary. In a hierarchic group one student is stronger than the other(s), can lead the interaction, and probably guide the other. In a complementary group the differences between the learners are such that members can either combine their capabilities to solve tasks, or teach each other what they know to expand their knowledge on a certain topic (e.g., Gutierrez-Santos et al., 2017).

Automated assessment

Recently, several tools have been designed to automate the process of formative assessment, particularly in mathematics education (e.g., Stacey & Wiliam, 2012). Formative assessment systems contribute to learning and instruction by giving detailed yet idiosyncratic feedback to learners, and providing teachers with an overview of the learning in the classroom. In the present study, we were interested in grouping based on students' performance in a particular mathematical task.

We used a formative assessment system to automate the process of assessing students' performance with the Seeing the Entire Picture (STEP) platform: a domain-specific formative assessment system developed in our laboratory (Olsher, Yerushalmy, & Chazan, 2016; Yerushalmy, Nagari-Haddif, & Olsher, 2017). With STEP, mathematical questions are posed to students using dedicated GeoGebra-based applets. Students solve the problem and submit their answers, for example, as a finite solution to a question, as a set of examples supporting or refuting a mathematical claim, or as examples of a mathematical idea. STEP collects the submitted answers and characterizes them based on their mathematical properties and on their correctness. The system then uses this elaborated data to report back to students (for further learning) and to the teacher (for further instruction). For example, students may be asked to choose three pairs of points and build three linear functions on these points using GeoGebra. When submitting answers, the system can provide information on the properties of the students' answers, such as whether the function increases or decreases, or whether one of the points chosen is on the Y axis. The system can also provide information on correctness, for example, indicating whether the functions generated passes through the selected points. Using this information, teachers can decide which students need further instruction, and of what type.

The answers submitted by the student through STEP in the above example of linear functions is a representation of the student's personal example space: a repertoire of available examples with regard to a certain mathematical idea (Sinclair, Watson, Zaskis, & Mason, 2011). A learning goal for individual students may be to extend their personal example space. One promising path for widening one's personal example space would be conducting a dialogue with a teacher, a peer, or perhaps an object.

The objective of our research project is to design, implement and test a support system for informed grouping. Therefore, the following study aims to unravel teachers' considerations, in grouping students using data obtained from STEP. We asked: (a) How do teachers use STEP data to group students? and (b) What were their considerations for grouping students?

METHODOLOGY

Population

Twenty graduate students in mathematics education, all of them mathematics school teachers at various levels, who attended a course on automated formative assessment in mathematics education, participated in stage 1 of the experiment. Twelve of the students participated also in stage 2. Participation was voluntarily.

The task

Students were presented with a GeoGebra-based applet (Figure 1) and given the following instructions: "Choose (red) points with the '*new points*' button to build a linear function whose graph passes through the points you have chosen. If you think it is not possible, explain why. Submit three examples of three different pairs of points." The problem was designed to probe the personal example space of learners: they can choose their pairs of points of their liking, and build linear functions on them. The presentations of pairs of points in the applet was only partially randomized: the probability of the applet presenting some cases was higher than that of presenting some other cases. For example, the case "one point is on the Y axis" appears in 25% of cases; the case "both points have the same Y value" appeared in 15% of cases.

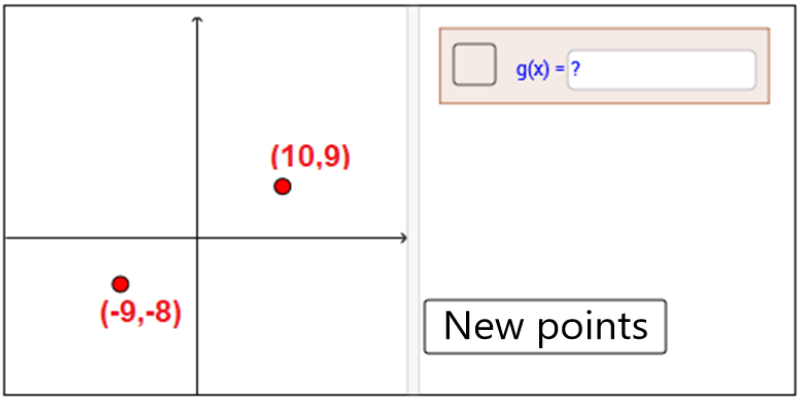

Figure 1: GeoGebra-based applet for the activity.

Procedure

<u>Stage 1</u>. The teachers solved the activity described above (Figure 1). Every answer contained three examples that represented their "personal learning space" (although their actual personal example space may be larger). Next, the teachers participated in a whole-class discussion of the various solutions to the problem submitted by the different participants.

Since teachers and students are the same individuals, but we need to distinguish between them, we will refer to "students" when discussing personal example spaces using pseudonyms (e.g., Arale, Yochi); and to "teachers" (e.g., Teacher a, Teacher g) when discussing the grouping being performed.

<u>Stage 2</u> was carried out a week later. Twelve teachers of the 20 decided to participate in this stage. Before the beginning of the lesson 12 teachers received printouts of 20 answers (i.e., the personal example spaces of the 20 students in stage 1). In addition, every teacher received an experiment form, in it a table with three columns and eight rows. At the headline of the rows printed "student A", "Student B" and "why were they paired?"

First, the teachers were asked to take five minutes to think of possible ways to group students based on these answers, without producing anything in writing. Second, the teachers participated in a 50-minute lecture and discussion about the underlying principles of grouping, the concept of a personal example space, and how it may support complementarity in grouping. Third, the teachers were asked to look at the printouts of the answers, presented under fictitious “student” names, and group the students by pairs, entering each pair in a table and providing the reason for the grouping: “Group the students based on their answers. The answers should be complementary. The objective is to expand the students' personal example space, assuming the two students will later learn together. Explain your choice.” Fourth, the teachers participated in a discussion in which they presented some of groupings and explained their rationale for them. The lecturer of the course did not participate in the experiment.

Data collection

Nine of the twelve teachers who agreed to participate in the experiment agreed to submit the experiment forms (the completed tables). The researchers were blind to the process of submission. The teachers submitted a total of 53 groupings (average 5.78). See Table XY for the number of groupings per teacher.

Data analysis

The first author created definitions for, and coded the data based on, three strategies for grouping: similarity, hierarchy, and complementarity. Based on these definitions, the second author coded 20% of the data, and discussed it with the first author, and the two identified in the data two sub-strategies of hierarchic grouping. The three authors then looked again at the definitions and the coding; together they reached a finer articulation of each strategy. Finally, the first author recoded all the data according to the resolution by the three authors.

FINDINGS

We start this section providing quantitative data about the distribution of groupings done by the teachers (see table 1), followed by an explanation of the way they manifest in the context of the given STEP activity and an example for how they were manifested. This findings section ends with an illustration of a case study that sheds light on the complexity of this grouping-task. Table XY lists the groupings submitted by the nine teachers, and the categories to which the groupings were assigned.

Table 1. Type and distribution of groupings by teachers

Teacher	**Number of groupings**	**Similarity**	**Hierarchy**		**Complementarity**
			PES Size	Correct answers	
Teacher a	**7**	2	3		2
Teacher b	**5**		1		4
Teacher c	**7**	1	1	1	4
Teacher d	**8**		5	2	1
Teacher e	**6**		5		1
Teacher f	**4**		1	1	2
Teacher g	**7**		2	3	3

Teacher h	**4**		4		
Teacher i	**5**	1	1	3	
Total	**53**	**4**	**23**	**10**	**17**

Table XY shows that the most popular grouping type (33 out of 53) was hierarchy-based. Hierarchy was the one strategy used by all teachers. Complementarity and similarity were used by fewer teachers, seven and three correspondingly. Teachers a and c have shown the most versatility by grouping based on all three grouping strategies.

Strategy A: Similarity

In four cases (7.5%), three of the teachers chose to group students based on similarity. According to this strategy, students who displayed few differences based on one or more of the analyzed dimensions were grouped together. The teachers who used this strategy would often propose an assignment for them as a group, to expand the personal example space of both students in the same direction. This can be accomplished by correcting a common incorrect answer or by asking the learners to expand their similar personal example spaces toward a certain dimension. For example, Teacher a explained why he proposed to group Haviva and Arale (all names are fictitious) together: “Both of them created increasing functions, so I would like them to sit together to expand their example spaces and create more examples that are different from each other.”

Strategy B: Hierarchy

In 33 cases (62.3%), the teachers chose to group students based on hierarchy. According to these strategies, the teacher considered the submission of one student to be superior to that of another, and aimed for the better student to teach the weaker one. We identified two types of such hierarchical relationships determined by the teachers: (a) Correctness (10 cases, 18.9%), when the solutions of one student were correct and those of the other were not, in which case the student with the correct answers was expected to teach the one with the incorrect answers. For example, Teacher g explained why she grouped students Yochi and Haviva together (figure 2): “Yochi solved correctly and with variance [between answers]. Haviva didn’t solve the exercises correctly, so it’s possible that the first will teach the second.” (b) Size of the personal example space (23 cases, 43.4%), when large differences were found between the personal example spaces of the students. A particular case may be one in which the personal example space of one student is fully included in that of another. For example, Teacher e and Teacher h grouped Haviva and Arale based on hierarchy. Teacher e explained: “Arale provided correct examples although all were very similar (increasing functions), but Haviva does not seem to be able to create all of the examples, so he can receive help from Haviva.”

Strategy C: Complementarity

In 17 cases (32.1%), seven of the teachers chose to group students based on complementarity. According to this strategy, grouping was based on some differences between the partners, so that each learner may learn from the other and that both may expand their personal example space. Grouping according to this strategy was based on various degrees of differentness. We identified two types of such relationships determined by the teachers. (a) Functions orientation: according to this grouping strategy, at least one possible orientation of a linear function (up, down, constant) was not present in the submissions of one student, but it was present in the submission of the other. (b) Choice of points: the criterion upon which grouping was made was related to the points chosen by the students—for example, solving for functions that have points on the axes or for functions that are built out of points in different quadrants.

Note that strategies (a) and (b) are rather similar, but the *explanations* for the grouping given by the teachers were either based on the outcome (function orientation) or on choice. We did not, however, distinguish between these strategies because they were rather close to one another. For example, Teacher f grouped Arale and Ayelet together, explaining that "Arale could teach Ayelet about the intersection [of the function] with the axes, and Ayelet could teach Arale about [choosing] a point on the X axis [and] an example of a constant function."

In some of the cases, the personal example spaces of the two students were complementary, but nevertheless the teacher saw the relationship between the two students as hierarchic. We categorized these cases as hierarchic, but noted the fact that automated grouping systems would have categorized these cases more accurately.

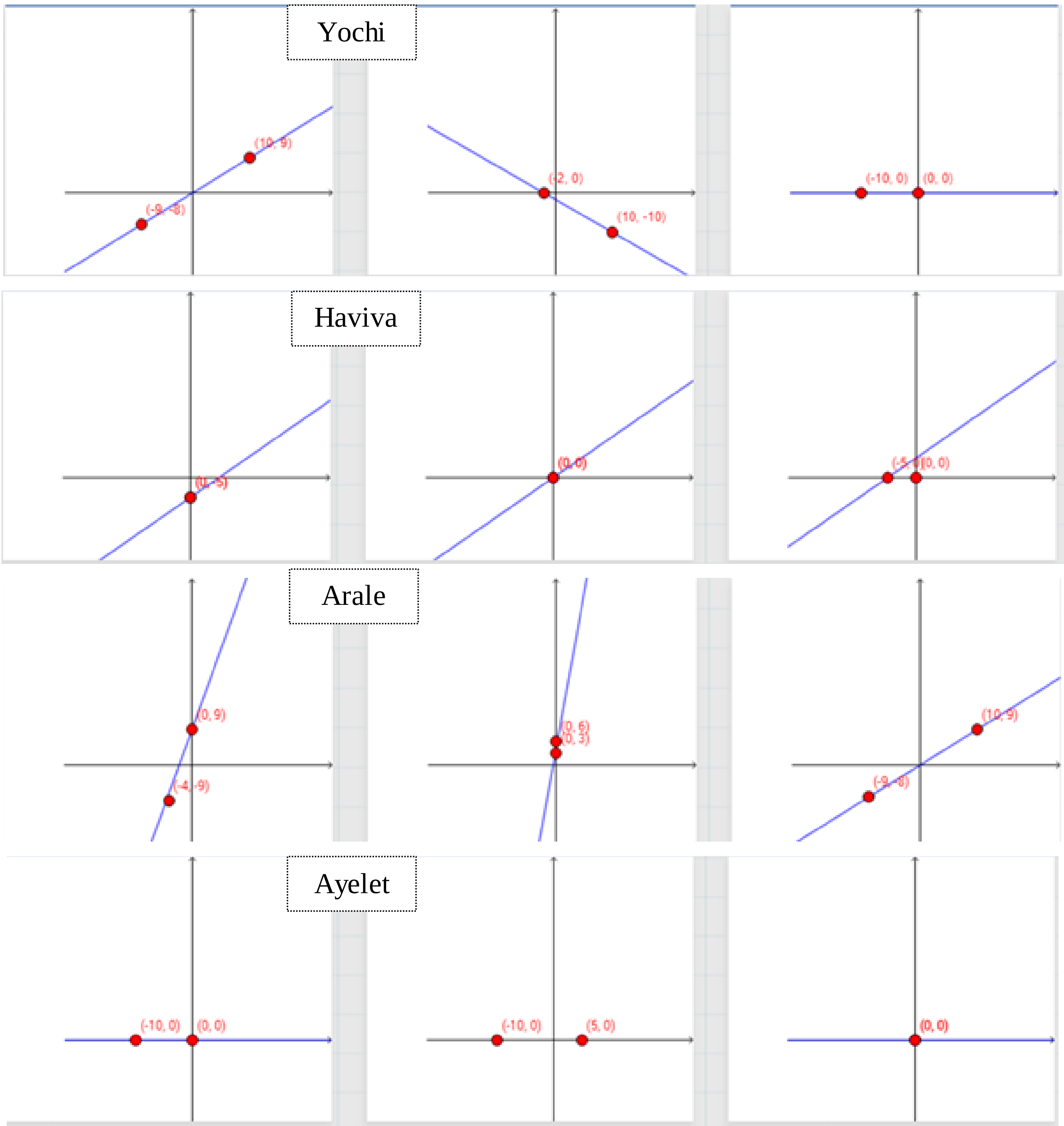

Figure 2. Four personal example spaces of four students.

Same grouping, different strategy

In nine cases, certain groupings were chosen by more than one teacher. In two instances, teachers' explanations were coded as the same strategy, and in seven instances as different strategies. In the examples above Haviva and Arale were grouped together by Teacher e and Teacher h based on hierarchic considerations, and by Teacher a based on similarity considerations. Five of the groupings appeared twice, three by three teachers and one by four teachers. One possible explanation for this unlikely finding is that five of the groupings were based on answers that were close physically in the printout of the answers.

DISCUSSION AND CONCLUSIONS

In a quest to use data from a mathematics-specific automated formative assessment system in order to perform informed groupings, we found the need to unravel teachers' ways of grouping students using such data. We asked, (a) How do teachers group learners into small groups based on STEP data? and (2) What are the teachers' considerations in making these groupings?

We produced three main findings. First, we made a new distinction between grouping strategies obtained from the literature and showed how they manifest in the context of a STEP-based activity. Second, we learned that despite the experimenters' request to group students in a way that each student will be able to teach the partner, with a planned bias toward complementarity, teachers grouped the students based mostly on other considerations, despite the fact that complementarity was present in the students' answers. This result appears to indicate an epistemological gap that teachers need to close. Most of the teachers were able to perform hierarchic grouping based on dimensions other than correctness, which suggests that grouping based on complementarity may be more complex. Third, we found that different teachers grouped students based on different considerations, suggesting that grouping by humans is not an objective process, and allows for diverse approaches.

These findings contribute to the hypothesis that informed grouping may be challenging for teachers, because it requires knowledge of the learning task, choosing dimensions for grouping, assessing learners based on these dimensions, and deciding how their performance in chosen areas should inform grouping. We argue that automated grouping systems and (mathematics) domain-specific automated formative assessment platforms such as STEP can together provide a satisfactory solution for informed grouping in mathematics. Processing large chunks of data in short time may overwhelm teachers, undermining their efforts to achieve data-informed groupings. Informed grouping could therefore benefit from automation. It could, just as well, free the teacher to address other aspects of learning, such as supporting the communication between the learners. As several computer scientists have shown, automated grouping algorithms can improve our ability to assess students' performance and provide informed suggestions for grouping (e.g., Abnar, Orooji, & Taghiyareh, 2012). In the case of STEP, informed grouping requires deep understanding of a students' personal example spaces, which STEP is already capable of achieving.

Complementary grouping helps avoiding the homogeneous vs. heterogeneous dichotomy. Complementarity is based on heterogeneity, that is, on difference, but it is also homogeneous because both learners could contribute to the learning of the other. With hierarchic grouping, one student may be categorized as inferior to another, so that one student can teach the other and help expand the other student's example space. In the case of complementarity, however, a student who may have displayed a narrower personal example space than the other student, may be still able to contribute to the work of the stronger student and help expand the stronger student's example space.

The following questions remain to be answered: Is complementary grouping superior to grouping based on similarity, hierarchy, or both? And if yes, what are the dimensions along which complementarity should be achieved with respect to students' personal example spaces? In our effort to automate the process of grouping, we sought to understand the ways in which teachers used STEP-based data to propose ways of grouping. We believe that the present study has contributed to creating an effective automated grouping systems using automated formative assessment systems.

REFERENCES

Abnar, S., Orooji, F., & Taghiyareh, F. (2012, January). An evolutionary algorithm for forming mixed groups of learners in web based collaborative learning environments. In *Technology Enhanced Education (ICTEE), 2012 IEEE International Conference on* (pp. 1-6). IEEE.

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi delta kappan*, *80*(2), 139-148.

Barron, B. (2003). When smart groups fail. *The journal of the learning sciences*, *12*(3), 307-359.

Connor, C. M., Morrison, F. J., Fishman, B., Crowe, E. C., Al Otaiba, S., & Schatschneider, C. (2013). A longitudinal cluster-randomized controlled study on the accumulating effects of individualized literacy instruction on students' reading from first through third grade. *Psychological Science*, 24(8), 1408-1419.

Dillenbourg, P. (1999). What do you mean by collaborative learning? *Collaborative learning: Cognitive and Computational Approaches*, 1-19.

Gutierrez-Santos, S., Mavrikis, M., Geraniou, E., & Poulovassilis, A. (2017). Similarity-Based Grouping to Support Teachers on Collaborative Activities in an Exploratory Mathematical Microworld. *IEEE Transactions on Emerging Topics in Computing*, *5*(1), 56-68.

Hoyles, C., Healy, L., & Sutherland, R. (1991). Patterns of discussion between pupil pairs in computer and non-computer environments. *Journal of Computer Assisted Learning*, *7*(4), 210-228.

Konert, J., Burlak, D., & Steinmetz, R. (2014). The group formation problem: an algorithmic approach to learning group formation. In *European Conference on Technology Enhanced Learning* (pp. 221-234). Springer, Cham.

Maqtary, N., Mohsen, A., & Bechkoum, K. (2017). Group Formation Techniques in Computer-Supported Collaborative Learning: A Systematic Literature Review. *Technology, Knowledge and Learning*, 1-22.

Olsher, S., Yerushalmy, M., & Chazan, D. (2016). How might the use of technology in formative assessment support changes in mathematics teaching? *For the Learning of Mathematics*, *36*(3), 11-18.

Schwarz, B. B., Neuman, Y., & Biezuner, S. (2000). Two wrongs may make a right... If they argue together!. *Cognition and instruction*, *18*(4), 461-494.

Sinclair, N., Watson, A., Zazkis, R., & Mason, J. (2011). The structuring of personal example spaces. *The Journal of Mathematical Behavior*, *30*(4), 291-303.

Stacey, K., & Wiliam, D. (2012). Technology and assessment in mathematics. In *Third international handbook of mathematics education* (pp. 721-751). Springer, New York, NY.

Wichmann, A., Hecking, T., Elson, M., Christmann, N., Herrmann, T., & Hoppe, H. U. (2016, August). Group Formation for Small-Group Learning: Are Heterogeneous Groups More Productive?. In *Proceedings of the 12th International Symposium on Open Collaboration* (p. 14). ACM.

Webb, N. M. (2009). The teacher's role in promoting collaborative dialogue in the classroom. *The British Journal of Educational Psychology*, *79* (1), 1–28.

Yerushalmy, M., Nagari-Haddif, G., & Olsher, S. (2017). Design of tasks for online assessment that supports understanding of students' conceptions. *ZDM*, *49*(5), 701-716.

DOI: 10.17185/duepublico/70750
URN: urn:nbn:de:hbz:464-20191115-150641-6