European Journal of Science and Mathematics Education, 2023, **11**(1), 15-32
ISSN: 2301-251X (Online)

OPEN ACCESS

Research Article

Differences in students' mathematics knowledge in homogeneous and heterogeneous groups

Boris Černilec [1]
0000-0003-2330-8123

Mara Cotič [2]
0000-0002-4958-4858

Darjo Felda [2]
0000-0002-8588-0654

Daniel Doz [2*]
0000-0002-6942-6937

[1] Zavod za Gluhe in Naglušne Ljubljana, Ljubljana, SLOVENIA
[2] Faculty of Education, University of Primorska, Koper, SLOVENIA
[*] Corresponding author: daniel.doz@upr.si

Citation: Černilec, B., Cotič, M., Felda, D., & Doz, D. (2023). Differences in students' mathematics knowledge in homogeneous and heterogeneous groups. *European Journal of Science and Mathematics Education, 11*(1), 15-32. https://doi.org/10.30935/scimath/12431

ARTICLE INFO

Received: 7 May 2022

Accepted: 16 Aug 2022

ABSTRACT

The question of grouping students into homogeneous and heterogeneous groups is not new, but it does not find an unambiguous answer in the literature, especially in mathematics. In this paper, we address the question of whether grouping students into homogeneous and heterogeneous groups in mathematics improves their knowledge. The quasi-experiment involved 126 8-grade (i.e., 13-14 years old) Slovenian primary school students, who were divided into two equal groups: the control group worked in homogeneous groups, and the experimental worked in heterogeneous groups. The results of the post-test show that the students from the experimental group had better results in mathematics, which indicates that heterogeneous groups should be preferred in mathematics. Lesson observations have identified differences in teacher behavior: educators working in homogeneous groups tend to give students with lower competencies tasks from lower taxonomic levels, and teachers favor abler groups. Such differences have not been observed among teachers working in heterogeneous groups. The implications for educators are also discussed.

Keywords: mathematics, homogeneous groups, heterogeneous groups, differentiation

INTRODUCTION

The goals of modern education are particularly demanding (Blanco Diez, 2018), thus there is a tendency to help each individual to acquire high-quality knowledge, in order to avoid being "left behind" (cf. Dee & Jacob, 2011). Consequently, effort should be made to adapt learning and teaching to the abilities of learners (Liu, 2007). An example of such individualization of learning is represented by working and studying in small groups, which might be their composition homogeneous or heterogeneous (Oetzel, 1998; Raftu, 2016; Wyman & Watson, 2020). In the literature, ample discussion regarded which way of grouping students is more suitable (cf. Esposito, 1973; Wyman & Watson, 2020); however, the documented benefits of grouping students regard a more rational use of time in managing the learning process, a better response to learners' questions with clearer explanations, listening to students and reacting to their responses, as well as monitoring the progress of each individual student within the group in a more efficient and easy way (cf. Schullery & Schullery, 2006).

Consequently, students are expected to have better performances (Strmčnik, 1993). Research has shown positive effects of both homogeneous and heterogeneous groupings (Shullery & Shullery, 2006).

The question of the effect of grouping pupils into homogeneous and heterogeneous groups has been explored also in mathematics. For instance, some researchers (Burris et al., 2006; Leonard, 2001) have proved that students from heterogeneous groups have better achievements on some standardized tests of mathematical knowledge, other researchers (Fuchs et al., 1998) proved that the contrary is true, and some studies (Huang, 2009) proved that there are no statistically significant differences in mathematics achievements between homogeneous and heterogeneous groups. Thus, it is still not clear which way of grouping is the most efficient for students to learn mathematics.

The aim of the present study is to contribute to the literature on the topic of grouping students into homogeneous and heterogeneous groups in mathematics. The nature of the present study is quasi-experimental: we wanted to investigate which group has the highest achievements in mathematics. Results are presented and suggestions for educators are discussed.

THEORETICAL FRAMEWORK

A lot of research has been done in the field of differences in knowledge between students of heterogeneous and homogeneous groups (cf. Burris et al., 2006; Esposito, 1973; Fuchs et al., 1998; Huang, 2009; Leonard, 2001; Oetzel, 1998; Raftu, 2016; Schullery & Schullery, 2006; Strmčnik, 1992; Wyman & Watson, 2020). However, we need to be careful when generalizing these results. Some research at this level of education shows that grouping students can be effective if teaching methods and learning materials are adapted to students' needs. For example, Askew and Wiliam (1995), in a review of various studies, found positive effects of grouping in mathematics in a higher-level group of students if the learning material was written and adapted for them. The research adds that it is not entirely clear what contributed to students' progress, whether the teaching material or the attention and professionalism of the teacher. Similar results are present in other research as well (Fox, 1979; Gregory, 1984; Kulik & Kulik, 1984), from which it can be concluded that we should not expect positive effects of grouping students on their achievements if teachers do not use appropriate teaching materials.

Some studies, which dealt with the grouping of students by ability in the first and second grades primary school, found that the grouping of students based on their abilities has no effect on students' outcomes (Adamič, 1996; Dawson, 1987; Slavin, 1987; 1990; Žagar, 2004). Pupils classified into homogeneous groups did neither better nor worse than the control group of students in heterogeneous classes. These findings might be applied also to gifted students, average-able students, and also to students with poorer abilities.

Some research made to compare mathematics achievements of students in homogeneous and heterogeneous groups found that grouping students into homogeneous groups did not affect math performance (Linchevski, 1995) or students in heterogeneous groups did usually better (Boaler, 1997; Leonard, 2001). Math achievements of students with lower abilities in the level groups were significantly lower than those of equally able students in heterogeneous groups since less able students cope better with tests in heterogeneous classes because both themselves and teachers have higher expectations when writing tests and grading (Linchevski, 1995; Venkatakrishnan & Willliam, 2003; Valenčič Zuljan et al., 2015; Žakelj et al., 2009). An additional explanation is that teachers in higher-level groups make greater use of teaching methods and forms that develop students' creativity and critical thinking, while lower-level groups use a transmission model of teaching to a greater extent (Page, 1992). Thus, teachers' behavior differs according to the level group in which they taught since teachers provide students with higher learning achievements with better learning opportunities, but they also influence the student's learning self-image and how peers look at each other (Ireson & Hallam, 2001). Furthermore, some educators teaching in homogeneous groups prefer more disciplinary problems (Ireson & Hallam, 2001).

Considering students' perspective, learners pointed out that a good understanding of the material in individual level groups often prevents them from working in a group, as it is too fast or too slow (Boaler, 1997). As an additional issue, students with lower learning achievement are more likely to be the target of underestimation and ridicule (Ireson & Hallam, 2001) and show lower self-confidence (DiMartino, 2005).

Additional weaknesses of grouping students include polarization, creating elitism, low expectations for lower-level students, and promoting segregation (DiMartino, 2005).

Based on all the mentioned research, especially at the upper level of primary school, we could conclude that there are no consistent and reliable results on the positive or negative effects of grouping students into homogeneous groups. The conflicting results of different research are often due to methodological difficulties (e.g., different criteria for classifying students into individual level groups; cf. Huang, 2009), difficulties in controlling the teaching methods used by teachers in individual groups, and difficulties in controlling teachers' views. These variables were not always adequately monitored in these studies. Nevertheless, the abovementioned studies evidenced that students from heterogeneous groups had better achievements.

EMPIRICAL RESEARCH

Aims of the Research

The aim of the present research is to verify the impact that grouping students into homogeneous and heterogeneous groups has on their knowledge of mathematics. In particular, with the present work, we want to show that working within a heterogeneous group enhances students' mathematical knowledge on all three taxonomic levels:

1. conceptual knowledge,
2. procedural knowledge, and
3. problem solving (Gagne, 1985).

Thus, we wanted to prove that students from the experimental group, composed of students working in heterogeneous groups (HET), have higher achievements than students from the control group, who worked in homogeneous groups (HOM).

Research Design and Methodology

The research design for this study was quasi-experimental. Specifically, we used a non-equivalent control-group design for the primary question of the study. The quasi-experimental research methodology is most appropriate in this case because the students were not randomly assigned to the experiment and control groups (Gall et al., 2007).

To prove the validity of the abovementioned research hypothesis, we applied the causal-experimental research method. Both quantitative and qualitative research methods were applied.

In forms of experimental research, the independent variable is manipulated. In this case, the independent variable is the method of grouping (homogeneous or heterogeneous). Two groups were considered in the present research, each one was taken from a Slovenian elementary school, i.e., the experimental and the control group. Students from the experimental group were selected from an elementary school that applies heterogeneous group learning, while the control group consisted of students from an elementary school with homogeneous groups. In the previous school year, students from both groups were part of heterogeneous groups.

The pedagogical experiment was deemed the most suitable to test the differences between the experimental and control group. Quantitative data were collected with two tests of mathematical knowledge. Qualitative data related to the quality of lessons were collected through observation, interviews with teachers, written and oral analysis of the lesson plans, and the analysis of the contents of students' notebooks. This permitted us to have a better understanding of the school practice. In particular, during classroom observations, the focus of the researchers was on the material used in homogeneous and heterogeneous classes, whether the lessons were equally as challenging in both grouping type classes, and whether the teachers had the same expectations from both grouping types. Lesson plans gave the researchers information about the content being taught to the homogeneous and heterogeneous classes and whether the rigor was comparable. Researchers observed the groups on multiple occasions, and both homogeneous and heterogeneous classes were observed. Field notes were written during all observations. The teachers provided weekly lesson plans during the research process. The collected data was then analyzed using coding.

Participants

Participants in the study were 126 students from 8th grade of two Slovenian elementary schools. The HET (i.e., experimental) group was comprised of 63 students from three different classes. The HOM (i.e., control) group was also composed of 63 students. Both the chosen students' schools were suburban, properly equipped, and with adequate didactical material and information-communication technology. Students' classes were big, there was enough illumination inside them, and the number of students in each of them was mainly equal (namely, from 20 to 23 students). In both schools, lessons occurred only in the mornings, thus students always had free-from-class afternoons.

Experiment

The quasi-experiment is a single factor one. Both the experimental and the control groups were determined on the basis of whole classes of the chosen elementary schools. This consequently means that before the experiment, no random selection of the participants from the classes was applied, which would have been nearly impossible in pedagogical practice (cf. Grabbe, 2015).

The experiment started in September 2018 and ended in September 2019. All participants were in the 8th grade of elementary school in the school year 2018-19. During this school year, students from the experimental group worked in heterogeneous groups, while students from the control group were working in homogeneous groups. During the year, several observations of the lessons were performed by the researchers. The experiment was concluded at the early beginning of the next school year when students were in 9th grade.

In order to internally validate the experiment, at the beginning of it, we controlled the most influential factors, such as students' grades in Slovenian language and mathematics, students' gender, and socioeconomic status (SES). Regarding students' grades in mathematics, we found that students from HET ($M=3.89$; $SD=1.02$) and HOM ($M=3.97$; $SD=1.03$) had similar achievements ($t(124)=-.435$; $p=.664$). On the other hand, students' grades in the Slovenian language in the HET ($M=3.71$; $SD=1.05$) group and the HOM ($M=4.17$; $SD=.814$) did differ significantly ($t(124)=-2.744$; $p=.007$).

Regarding the SES, we computed it considering parents' education, the possibility of using the computer at home, the possibility of looking up words in the dictionary, the number of books at home, and the presence of the student's desk to write homework and study (see **Appendix A**; cf. Østbø & Zachrisson, 2021; Saifi & Mehmood, 2011). With the χ^2-test we check for differences between the HOM and HET groups, finding no statistically significant difference considering parents' education ($\chi^2(3)=3.403$; $p=.334$), the possibility of using the computer at home ($\chi^2(1)=1.211$; $p=.271$), the possibility of looking up for words in the dictionary ($\chi^2(1)=1.201$; $p=.273$), in the number of books ($\chi^2(3)=1.676$; $p=.642$), and the presence of the student's desk ($\chi^2(1)=.000$; $p=1.000$).

There were seven teachers teaching the HOM and HET groups; their level of instruction (all with university education), age ($M_{HOM}=36$; $M_{HET}=42$), and years of service ($M_{HOM}=12.6$; $M_{HET}=13.1$) were very similar.

Materials

Students' received a pre- and post-test of mathematical knowledge, which was made by the authors. The psychometric characteristics of the test were initially tested on a sample of randomly chosen 100 grade-8 students from five elementary schools in the Central Slovenian region. The test was composed of questions, which encompassed the topics included in the national program of mathematics (Učni Načrt, 2011), and we considered the objectives and standards of knowledge described in it. There were 21 tasks in each of the tests of knowledge.

The pre-test was composed of seven parts, as follows:

1. sum and subtraction of rational numbers,
2. polygons,
3. multiplication and division of rational numbers,
4. powers,
5. relationships between variables,

6. circle and circumference, and
7. computations with variables (i.e., algebra).

The post-test was composed of questions regarding the Pythagorean theorem. Both the pre- and post-test included exercises from all three of Gagne's (1985) taxonomic levels. Each exercise was evaluated on a scale from 0 to 3 points; thus, the whole test was worth a maximum of 63 points. Students had two school hours (i.e., one hour and 30 minutes) to solve each of the tests.

In order to guarantee the validity of the test, we made the tests considering the national program for Mathematics for primary schools (Učni Načrt, 2011). Translated copies of the tests are present in **Appendix B**. When deciding the goals and standards that were assessed, we considered Gagne's (1985) taxonomic levels:

1. Conceptual knowledge,
2. Procedural knowledge, and
3. Problem solving.

30% of the tasks in the test regarded conceptual knowledge (five tasks); 55% the procedural knowledge (13 tasks); 15% of the tasks were in the highest taxonomic level (three tasks; cf. RIC, 2005). The content validity of the tests was assessed by the seven independent teachers.

The reliability of the test was checked using Cronbach's alpha coefficient. The pre-test had a reliability of α=.895, while the post-test had a reliability of α=.899. Both reliabilities are very good.

Data Analysis

All data were analyzed with the SPSS statistical program. Since data was not deemed suitable for parametric statistical tests (i.e., the violation of the assumptions of normality and equality of variances, the categorical nature of the variables), non-parametric were preferred. In particular, we used the Mann-Whitney U-test to check the differences between the HOM and HET groups, and the Wilcoxon signed rank test to check for possible differences in the pre- and post-test. The Mann-Whitney U-test implies the calculation of *the U-stat*istic for each group we want to compare. For each group, i.e., 1 and 2, we compute:

$$U_1 = n \cdot n_1 + \left(\frac{n_1(n_1+1)}{2}\right) - R_1,$$

$$U_2 = n \cdot n_2 + \left(\frac{n_2(n_2+1)}{2}\right) - R_2,$$

where n_1 is the sample size of the first group, and n_2 is the sample size of the second one, while R_1 is the sum of ranks assigned to the first group, and R_2 is the sum of ranks assigned to the second group. The null hypothesis, i.e., that the samples have the same median, is rejected if the *p*-value of the minimum between U_1 and U_2 (i.e., $U = \min\{U_1, U_2\}$) is smaller than the threshold α (Nachar, 2008). The minimum of U_1 and U_2 is called the *U*-statistic (Tallarida & Murray, 1987).

The Wilcoxon signed rank test was used to compare students' pre- and post-test results. The Wilcoxon test requires the computation of the difference between two paired variables, i.e., $W_i = X_i - Y_i$, and converting the paired sample to a one-sample test. The absolute value of W_i is computed and sorted using rankings R_i. Ranks R_i are then "signed", i.e., multiplied by +1 if W_i is positive, and by -1 if W_i is negative, $\operatorname{sgn}(W_i)R_i$. Let us denote with W_+ the sum of positive ranks and with W_- the sum of negative ranks. The Wilcoxon W is the smallest between W_+ and W_-. The null hypothesis, i.e., that the paired samples have the same median, is rejected if the *p*-value of W is smaller than the threshold α. The rank-biserial coefficients for the Mann-Whitney U-test (r_{MW}) and the Wilcoxon W test (r_W) have been computed, in order to measure the effect size with the following formulae (Kerby, 2014; Wendt, 1972):

$$r_{MW} = 1 - \frac{2U}{n_1 n_2},$$

$$r_W = \frac{4|T - (W_+ + W_-)/2|}{n_i(n_i+1)},$$

where $T = \sum_{j=1}^{n_1} \operatorname{sgn}(W_j)R_j$. Effects sizes (in module) of .30 are considered middle, and all those above .50 (in module) are considered big (LeBlanc & Cox, 2017).

Table 1. Descriptive statistics of the pre-test

	Group	n	Mean	Standard deviation	Median
Overall test	HOM	63	40.8	8.16	42
	HET	63	33.1	12.2	33
Taxonomy					
Conceptual knowledge	HOM	63	9.27	4.15	9
	HET	63	6.67	3.22	6
Procedural knowledge	HOM	63	26.22	7.91	27
	HET	63	21.86	10.96	23
Problem-solving	HOM	63	5.33	2.67	6
	HET	63	4.62	2.76	5

Table 2. The ranking of the two groups among the three Gagne's (1985) taxonomic levels in the pre-test

Topic	Group	M	SD	Difference	t-test		
					t	df	p
Sum and subtraction of rational numbers	HOM	6.94	1.93	.59	-1.840	121.1	.068
	HET	6.35	1.65				
Polygons	HOM	5.98	2.10	1.08	-2.601	119.8	.004**
	HET	4.90	2.54				
Multiplication and division of rational numbers	HOM	5.86	2.16	1.10	-2.910	124	.004**
	HET	4.76	2.06				
Powers	HOM	6.11	1.62	.41	-1.404	124	.163
	HET	5.70	1.68				
Relationships between variables	HOM	7.37	1.78	1.21	-3.674	124	<.001***
	HET	6.16	1.90				
Circle and circumference	HOM	5.52	2.35	.77	-1.708	121.2	.090
	HET	4.75	2.74				
Algebra	HOM	4.02	2.77	.08	.160	124	.873
	HET	4.10	2.78				

Note. **p<.05 & ***p<.001

RESULTS

Analysis of the Pre-Test

The initial differences between the HET (experimental) and HOM (control) group were tested with the pre-test. The test, as described earlier, comprised 21 questions regarding geometry and arithmetic (together with algebra). All three Gagne's (1985) taxonomic levels were included. In **Table 1**, we present the descriptive statistics of students from the HOM and HET groups, divided among the three Gagne's (1985) levels.

With the Mann-Whitney U-test, we checked for possible differences between the HOM and HET groups among the three taxonomic levels. We found statistically significant differences between the control and experimental group on the level of conceptual (U=1,265; p<.001; r=.36) and procedural knowledge (U=1,562; p=.039; r=.21), however there were no statistically significant differences between the groups concerning the problem-solving taxonomic level (U=1,678; p=.133; r=.15). Thus, students from the control group (HOM) had higher achievements on the first two Gagne's (1985) taxonomic levels than students from the experimental group (HET).

Additional information about the achievements in each part of the test is present in **Table 2**. Students from the control group (HOM) had better achievements in polygons, multiplication, and division of rational numbers, and relationships between variables. Concerning the remaining topics of the test, no statistically significant differences were found.

Analysis of the Post-Test

The final differences between the HET and HOM groups were tested with the post-test. The test, as described earlier, comprised 21 questions regarding geometry (Pythagorean theorem). All three of Gagne's (1985) taxonomic levels were included. In **Table 3**, we present the descriptive statistics of the post-test, divided among the three Gagne's (1985) levels.

Table 3. Descriptive statistics of the post-test

	Group	n	Mean	Standard deviation	Median
Overall test	HOM	63	34.0	8.95	34
	HET	63	42.7	7.22	42
Taxonomy					
Conceptual knowledge	HOM	63	6.51	4.58	6
	HET	63	6.75	3.85	6
Procedural knowledge	HOM	63	23.33	8.41	22
	HET	63	30.30	5.86	31
Problem-solving	HOM	63	4.14	2.54	4
	HET	63	5.65	2.48	6

Table 4. The Wilcoxon signed-rank test

	Group					
	HET			HOM		
	W	p	r	W	p	r
Overall test	46	.000	-.95	1840	.000	.95
Taxonomy						
Conceptual knowledge	550	.885	-.00	1504	.000	.70
Procedural knowledge	62.5	.000	-.93	1520	.000	.78
Problem-solving	408.5	.000	-.51	1329	.000	.61

The Mann-Whitney test has shown that there are no statistically significant differences in what concerns conceptual knowledge: both the HOM and HET group had similar achievements (U=1,851; p=.515; r=.067). However, students from the HET group had statistically better achievements than students from the HOM group, both concerning procedural knowledge (U=1,027; p<.001; r=.48) and problem-solving (U=1,320; p=.001; r=.34).

Differences in Pre- and Post-Test

In **Table 4**, the results of the Wilcoxon signed rank test to compare students' pre- and post-test scores are presented. We found that students from the experimental group (HET) had better achievements in procedural knowledge (p<.001) and problem-solving (p<.001), while there were no significant differences in conceptual knowledge (p=.885). On the other hand, students from the control group (HOM) had an overall decrease in achievements (p<.001) at all taxonomic levels. The overall performance on the post-test from the HET group was better than the performance on the pre-test (p<.001), while the performance of the HOM group was overall worse than the one in the pre-test (p<.001).

Qualitative Analysis and Interpretation

In the present research, we also observed lessons in all homogeneous and heterogeneous groups. The findings of the observation were used in the interpretation of the results, but they cannot be generalized. Lesson planning was analyzed by reviewing the preparations for the lessons made by the teachers and by reviewing the individual notebooks of the students. Teachers' lesson plans in both groups of teachers did not differ and were consistent with the National program (Učni Načrt, 2011). In the study of teaching preparations and observation of mathematics lessons, we noticed mainly differences in:

1. internal differentiation and individualization during mathematics lessons in HET and HOM groups, and
2. treatment of mathematical content according to Gagne's (1985) taxonomic levels.

Based on the observation of the lessons of the control group in smaller homogeneous and heterogeneous groups, we can conclude that all mathematics teachers in the eighth grade participated in lesson preparation, regardless of which group they taught. Depending on the teaching in individual classes, the same teachers teach in different homogeneous groups based on knowledge and the classification of children into smaller level groups. Between homogeneous and heterogeneous smaller groups, there are noticeable differences in the planning of goals and standards of knowledge, as well as contents and activities in preparation for lessons. In the lower-level groups (i.e., in groups with lower-achieving students), the main emphasis is on conceptual and procedural knowledge. Problem-solving due to the poor ability of students is almost non-existent.

Conceptual knowledge, procedural knowledge and problem-solving are present only in the second- and third-level groups (i.e., among better-achieving students), but it appears most often in the highest-level group, occasionally also based on situations from everyday life. Occasionally, the lessons are not challenging enough for the students, as the vast majority of the more successful students get bored. Also, when better-achieving students gave incorrect answers, teachers coaxed them to develop correct answers, while lower-achieving students who were incorrect were usually ignored. Thus, despite the teachers' effort to use the optimal instructional method for each category of student, teachers' behavior was impacted by students' level.

Observation of classes in the experimental group showed that success, efficiency, organization of lessons, quality of lessons and climate in the department most often depend on the teacher. The classes emphasize conceptual and procedural knowledge; however, they do also face several problem-solving exercises from everyday life. Occasionally, more successful students did not receive additional activities and incentives, which is why disciplinary problems (i.e., minor misbehavior) also appeared. The teacher did often adjust the explanation of the learning material according to the individual student's ability. Individual work prevails in the consolidation of the learnt topics. In HET groups, both higher- and lower-achieving students were corrected by the teacher when they gave an incorrect answer to the question and no student was ignored when giving an incorrect answer.

When introducing and dealing with new learning material, the frontal teaching format prevailed in both the control and experimental groups. Among the teaching methods in the phase of introducing and dealing with new material, explanation dominates in both the HOM and HET groups, followed by the interview method and the method of written and graphic products. When consolidating, the most common method is written works. During the lesson, the teacher is the most noticeable, because most of the time he/she explains or explains individual solutions to the students.

From these observations, we might notice that students from HOM groups face fewer problem-solving exercises than students in HET groups, mainly because this taxonomic level is harder for lower-achieving students. Thus, these students face only routine-procedural tasks and are not engaged in more difficult and engaging problems. Moreover, some better-achieving students from HET groups took advantage of the situation and engage in minor misbehavior, however, were motivated and worked hard.

DISCUSSION

A lot of effort has been made in the literature to elucidate whether it is better to group students into heterogeneous and homogeneous groups (cf. Burris et al., 2006; Esposito, 1973; Fuchs et al., 1998; Huang, 2009; Leonard, 2001; Oetzel, 1998; Raftu, 2016; Schullery & Schullery, 2006; Strmčnik, 1993; Wyman & Watson, 2020). Results in the literature are ambivalent since some studies found no significant difference between homogeneous and heterogeneous groups (Adamič, 1996; Dawson, 1987; Slavin, 1987; 1990; Žagar, 2004), while others found a positive gain in heterogeneous groups (Boaler, 1997; Leonard, 2001). Therefore, the aim of the present work was to further explore the effect that grouping students into homogeneous and heterogeneous groups has on their math achievement. We conducted a quasi-experiment using two groups with equal sizes ($n=63$): the experimental group worked in heterogeneous groups (HET), while the control group worked in homogeneous groups (HOM). The purpose of the research was to prove the hypothesis that the experimental group would have significantly better achievements than the control group.

Before the experiment, background variables and students' knowledge were controlled. Students from both groups had a similar socioeconomic factor ($p>.05$) and similar grades in mathematics. The control group had a higher mean grade in the Slovenian language. Educators teaching in both the experimental and control group had a similar university education and years of service. Results of the 21-item pre-test have shown a difference between the groups, in favor of the control group regarding conceptual and procedural knowledge. No difference between groups has been found concerning problem-solving. The initial situation thus has shown a slight advantage of the control group concerning two taxonomic levels.

Considering the post-test, no differences have been noted between the groups with regard the conceptual knowledge. However, the experimental group had significantly higher achievements in procedural knowledge and problem-solving. Analyzing the differences between the pre- and post-test, results show that the experimental group has shown significant improvements in both procedural knowledge and problem-solving, while no significant differences in pre- and post-test conceptual knowledge have been detected. Students from the control group had lower achievements at all taxonomic levels.

The results of the experiment confirmed our main hypothesis, i.e., that students from heterogeneous groups would have better achievements than students in heterogeneous groups. Our findings are in accordance with the literature (Boaler, 1997; Leonard, 2001). Contrary to some studies that did not find any significant difference in students' performance when grouped in homogenous groups (Linchevski, 1995), we found that students from the control group had worse performance on the post-test.

Considering class observation, we might notice that students in HOM groups were less engaged in problem-solving activities (cf. Ireson & Hallam, 2001; Page, 1992), which were solved only by higher-achieving students. In the first level of HOM (the least able students), no problem-solving lessons occurred. The students did not encounter problem situations, but mainly learned concepts without their understanding and simple procedural knowledge. Teachers completely mistakenly think that these students are not capable of solving problems, which is completely wrong. They just have to face problems that are appropriate for their abilities (cf. Dee & Jacob, 2011). In the second-level group, the teaching was even more comparable to the teaching in heterogeneous groups, but there was a lack of problem-based knowledge. In the third-level group, the lessons were also problem-oriented. The teacher encouraged students to think creatively, critically, analytically, and systematically in solving mathematical problems. They discussed the ideas of the solution, the approaches, and the solutions obtained, and the teacher guided and listened to them, accepting, and considering their diversity.

Not having the possibility to face problem-solving exercises might have impacted students' learning, especially in lower-achieving ability-based groups (cf. DiMartino, 2005). On the other hand, students in HET groups dealt also with problems from the highest Gagne's (1985) taxonomic level, thus developing problem-solving skills.

In teaching and learning mathematics, it is imperative that levels are connected and equally represented (Rahbarnia et al., 2014). It is also wrong to think that levels are hierarchical: often understanding concepts is more mentally challenging than solving a problem (Denton, 2017). If we want the student to acquire mathematical knowledge and its application in life, he/she must know and understand the concepts and facts, he/she must have a good command of procedures and strategies for solving problems. When teaching in homogeneous groups at different levels, teachers teach very differently.

Moreover, another possible reason for the inefficiency of grouping students into homogeneous groups was observed in the belief of teachers that they teach a homogeneous group of students, so they do not need to differentiate and individualize work in individual level groups, although differences between students in individual groups exist and should be considered, also within individual groups (Boaler, 1997). As found during class observations, in heterogeneous groups, internal differentiation and individualization were more common.

In the classroom, the teacher must recognize and consider individual differences between students, find organizational models and implement the highest quality teaching so that students' knowledge will be lasting and of high quality. All students in a particular group are exposed to relatively equal learning difficulties and procedures. Individualization must consider and satisfy individual learning and other differences not only of the group but of the individual (Strmčnik, 2001). Differentiation arises from differences between students or from the different treatment of different children, and individualization comes from the individual or from differences in the child (Galeša, 1995).

Additionally, teachers' expectations of low-achieving students in HET groups might have contributed to their success (Linchevski, 1995; Venkatakrishnan & Willliam, 2003; Valenčič Zuljan et al., 2015; Žakelj et al., 2009). On the contrary, in HOM groups, teachers' behavior differs according to the level group in which they taught since teachers provide students with higher learning achievements with better learning opportunities, but they also influence the student's learning self-image and how peers look at each other (Ireson & Hallam, 2001; Page, 1992).

This research is subject to some limitations. Firstly, despite quasi-experiments being common in pedagogical research, randomization is needed to extend results to the population and generalize the findings. Moreover, although the used tests of mathematics knowledge have been validated, future studies might explore differences between HOM and HET using standardized tests of students' math knowledge (e.g., the PISA or TIMSS assessments). Additionally, the pre- and post-test applied in the present research had different contents; the pre-test assessed students' math knowledge of different topics, while the post-test assessed students' knowledge of the Pythagorean theorem. Future research might attempt to overcome this limitation. Concerning the qualitative part of the research, one of the limitations of the study concerns class observations since being in a position to critique a teacher's pedagogical practices could be an uncomfortable position for the teacher, thus influencing his/her teaching practice. Furthermore, in the present research, we investigated the differences in mathematical knowledge between homogeneous and heterogeneous groups. However, other social, personal, and behavioral characteristics of the students are equally important in the classroom. Hence, future research might try to explore the abovementioned factors as well.

The results of this research must also be interpreted with caution because the grouping of students into homogeneous groups influences the opinions and behaviors of teachers as well. In the qualitative part of the research, we have in fact highlighted how teachers who teach in homogeneous groups tend to prefer groups of more skilled students, to whom they assign exercises with higher taxonomic levels. On the other hand, groups comprised of less skilled students mostly solved exercises of lower taxonomic levels. Future research should also control and investigate the teacher variable, which might have a non-negligible impact on students' performance in both homogeneous and heterogeneous groups.

CONCLUSIONS

With the present research we wanted to investigate whether it was preferable to use homogeneous or heterogeneous groups during mathematics lessons. Although the literature on this subject shows conflicting results, some research has evaluated the positive aspects of grouping students into heterogeneous groups. The present study has confirmed this hypothesis. Furthermore, through qualitative research–i.e., observations of the lessons–it was noticed that the teachers of the homogeneous groups preferred to give students of lower levels exercises from a lower Gagne's (1985) taxonomy. In the lessons of the heterogeneous groups, these differences were not observed. Although it is therefore necessary to interpret these data with caution and not generalize them to the entire population, the present work aims to stimulate the debate regarding the possible use of heterogeneous groups in mathematics lessons.

Author contributions: BČ: data collection; **BČ & MC**: study conception and design; **MC & DF:** supervision; **BČ, MC, DF, & DD:** analysis and interpretation of results; & **DD:** draft manuscript preparation and references. All authors approve final version of the article.

Funding: The authors received no financial support for the research and/or authorship of this article.

Declaration of interest: Authors declare no competing interest.

Data availability: Data generated or analyzed during this study are available from the authors on request.

REFERENCES

Adamič, M. (1996). Model sukcesivnega kombiniranja temeljnega in nivojskega pouka ter dosežki nivojskih skupin [Model of successive combination of basic and level education and achievements of level groups]. *Sodobna Pedagogika* [*Modern Pedagogy*], *47*(1-2), 39-48.

Askew, M., & Wiliam, D. (1995). *Recent research in mathematics education.* HMSO/Osted.

Blanco Diez, J. C. (2018). *Learning contexts available for Japanese teachers in a top tier public high school: Encompassing a demanding work environment with adult education needs*. https://www.diva-portal.org/smash/get/diva2:1222725/FULLTEXT01.pdf

Boaler, J. (1997). Setting, social class and survival of the quickest. *British Educational Research Journal*, *23*(5), 575-596. https://doi.org/10.1080/0141192970230503

Burris, C. C., Heubert, J. P., & Levin, H. M. (2006). Accelerating mathematics achievement using heterogeneous grouping. *American Educational Research Journal, 43*(1), 137-154. https://doi.org/10.3102/00028312043001105

Dawson, M. M. (1987). Beyond ability grouping: A review of the effectiveness of ability grouping and its alternatives. *School Psychology Review, 16*(3), 348-369. https://doi.org/10.1080/02796015.1987.12085298

Dee, T. S., & Jacob, B. (2011). The impact of no child left behind on student achievement. *Journal of Policy Analysis and Management, 30*(3), 418-446. https://doi.org/10.1002/pam.20586

Denton, J. (2017). *Working with the IMPaCT taxonomy: Encouraging deep and varied questioning in the mathematics classroom* [PhD thesis, University of Warwick].

DiMartino, J. (2005). Reaching real equity in schools. *Education Digest, 70*(5), 9-13.

Esposito, D. (1973). Homogeneous and heterogeneous ability grouping: Principal findings and implications for evaluating and designing more effective educational environments. *Review of Educational Research, 43*(2), 163-179. https://doi.org/10.3102/00346543043002163

Fox, L. H. (1979). Programs for the gifted and talented. In: A. H. Passow (Ed.), *The gifted and talented: Their education and development* (pp. 104-126). University of Chicago Press.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Karns, K. (1998). High-achieving students' interactions and performance on complex mathematical tasks as a function of homogeneous and heterogeneous pairings. *American Educational Research Journal, 35*(2), 227-267. https://doi.org/10.3102/00028312035002227

Gagne, R. M. (1985). *The conditions of learning and theory of instruction.* Holt, Rinehart & Winston.

Galeša, M. (1995). *Specialna metodika didaktike* [*Special methodology of didactics*]. Didakta.

Gall, M. D., Gall, J. P., & Borg, W. R. (2007). *Educational research: An introduction*. Pearson.

Grabbe, J. W. (2015). Implications of experimental versus quasi-experimental designs. In: K. D. Strang (Ed.), *The Palgrave handbook of research design in business and management* (pp. 141-152). Palgrave Macmillan. https://doi.org/10.1057/9781137484956_10

Gregory, R. P. (1984). Streaming, setting and mixed ability grouping in primary and secondary schools: Some research findings. *Educational Studies, 10*(3), 209-226. https://doi.org/10.1080/0305569840100302

Huang, M. H. (2009). Classroom homogeneity and the distribution of student math performance: A country-level fixed-effects analysis. *Social Science Research, 38*(4), 781-791. https://doi.org/10.1016/j.ssresearch.2009.05.001

Ireson, J., & Hallam, S. (2001). *Ability grouping in education*. Paul Chapman Publishing. https://doi.org/10.4135/9781446221020

Kerby, D. S. (2014). The simple difference formula: An approach to teaching nonparametric correlation. *Comprehensive Psychology, 3*, 11-IT. https://doi.org/10.2466/11.IT.3.1

Kulik, C. L., & Kulik, J. (1984). *Effects of ability grouping on elementary school pupils: A meta-analysis* [Paper presentation]. The Annual Meeting of the American Psychological Association.

LeBlanc, V., & Cox, M. A. (2017). Interpretation of the point-biserial correlation coefficient in the context of a school examination. *The Quantitative Methods for Psychology, 13*(1), 46-56. https://doi.org/10.20982/tqmp.13.1.p046

Leonard, J. (2001). How group composition influenced the achievement of sixth-grade mathematics students. *Mathematical Thinking and Learning, 3*(2-3), 175-200. https://doi.org/10.1080/10986065.2001.9679972

Linchevski, L. (1995). Tell me who your classmates are and I will tell you what you are learning. *PME, XIX*(3), 240-247.

Liu, F. (2007). Personalized learning using adapted content modality design for science students. In *Proceedings of the 14th European Conference on Cognitive Ergonomics: Invent! Explore!* (pp. 293-296). https://doi.org/10.1145/1362550.1362612

Nachar, N. (2008). The Mann-Whitney U: A test for assessing whether two independent samples come from the same distribution. *Tutorials in Quantitative Methods for Psychology, 4*(1), 13-20. https://doi.org/10.20982/tqmp.04.1.p013

Oetzel, J. G. (1998). Explaining individual communication processes in homogeneous and heterogeneous groups through individualism-collectivism and self-construal. *Human Communication Research, 25*(2), 202-224. https://doi.org/10.1111/j.1468-2958.1998.tb00443.x

Østbø, I. U., & Zachrisson, H. D. (2021). Student motivation and parental attitude as mediators for SES effects on mathematics achievement: Evidence from Norway in TIMSS 2015. *Scandinavian Journal of Educational Research,* 1-16. https://doi.org/10.1080/00313831.2021.1939138

Page, R. (1992). *Lower track classroom: A curricular and cultural perspective*. Teachers College Press.

Raftu, G. (2016). Methods and techniques of instruction individualization and differentiation. Learning through cooperation or group work. *Bulletin of the Transilvania University of Braşov, Series VII: Social Sciences and Law, 9*(1-Suppl), 83-90.

Rahbarnia, F., Hamedian, S., & Radmehr, F. (2014). A study on the relationship between multiple Intelligences and mathematical problem solving based on revised Bloom taxonomy. *Journal of Interdisciplinary Mathematics, 17*(2), 109-134. https://doi.org/10.1080/09720502.2013.842044

RIC. (2005). *RIC*. https://www.ric.si/mma/izhodi%C5%A1%C4%8Da%20npz%20v%20o%C5%A1/20060706 11531042/

Saifi, S., & Mehmood, T. (2011). Effects of socioeconomic status on students achievement. *International Journal of Social Sciences and Education, 1*(2), 119-128.

Schullery, N. M., & Schullery, S. E. (2006). Are heterogeneous or homogeneous groups more beneficial to students? *Journal of Management Education, 30*(4), 542-556. https://doi.org/10.1177/1052562905277305

Slavin, E. R. (1990). Achievement effects of ability grouping in elementary and secondary schools: A best-evidence synthesis*. Review of Educational Research*, *60*(3), 471-499. https://doi.org/10.3102/00346543060003471

Slavin, R. E. (1987). Ability grouping and student achievement in elementary schools: A best-evidence synthesis. *Review of Educational Research, 57*(3), 293-336. https://doi.org/10.3102/00346543057003293

Strmčnik, F. (1992). *Problemski pouk v teoriji in praksi* [*Problem-based lessons in theory and practice*]. Didakta.

Strmčnik, F. (1993). *Učna diferenciacija in individualizacija v naši osnovni šoli* [*Learning differentiation and individualization in our elementary school*]. Zavod Republike Slovenije za šolstvo [Institute of the Republic of Slovenia for Education].

Strmčnik, F. (2001). *Didaktika. Osrednje teoretične teme* [*Didactics. Central theoretical topics*]. Znanstveni inštitut Filozofske fakultete [Scientific Institute of the Faculty of Arts].

Tallarida, R. J., & Murray, R. B. (1987). Mann-Whitney test. In R. J. Tallarida, & R. B. Murray (Eds.), *Manual of pharmacologic calculations* (pp. 149-153). Springer. https://doi.org/10.1007/978-1-4612-4974-0_46

Učni Načrt. (2011). *Učni Načrt*. https://www.gov.si/assets/ministrstva/MIZS/Dokumenti/Osnovna-sola/Ucni-nacrti/obvezni/UN_matematika.pdf

Valenčič Zuljan, M., Cotič, M., Felda, D., Magajna, Z., & Žakelj, A. (2015). *The efficiency of homogeneous and heterogeneous grouping of students in mathematics*. Verlag Dr. Kovač.

Venkatakrishnan, H., & William, D. (2003). Tracking and mixed ability grouping in secondary school mathematics classrooms: A case study. *British Educational Research Journal*, *29*(2), 189-204. https://doi.org/10.1080/0141192032000060939

Wendt, H. W. (1972). Dealing with a common problem in social science: A simplified rank-biserial coefficient of correlation based on the statistic. *European Journal of Social Psychology, 2*(4), 463-465. https://doi.org/10.1002/ejsp.2420020412

Wyman, P. J., & Watson, S. B. (2020). Academic achievement with cooperative learning using homogeneous and heterogeneous groups. *School Science and Mathematics, 120*(6), 356-363. https://doi.org/10.1111/ssm.12427

Žagar, D. (2004). Nivojski pouk v devetletni osnovni šoli [Level lessons in a nine-year primary school]. *Šolsko Polje* [*School Field*], *15*(5/6), 29-51.

Žakelj, A., Cankar, G., Bečaj, J., Dražumerič, S., Kern, J., & Rosc Leskovec, D. (2009). *Povezanost rezultatov pri nacionalnem preverjanju znanja s socialno-ekonomskim statusom učencev, poukom in domačimi nalogami*. Poročila o raziskavi [*Correlation of national test scores with students' socioeconomic status, lessons, and homework. Research reports*]. Zavod Republike Slovenije za šolstvo [Institute of the Republic of Slovenia for Education].

APPENDIX A

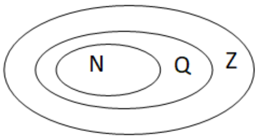
1. What is the highest level of schooling completed by your parents?
    a. University
    b. Secondary school
    c. Elementary school
    d. I do not know
2. Which of the following are in your home?
    a. A desk to study at
    b. A computer you can use for schoolwork
    c. A dictionary
    d. Books
3. How many books are there in your home?
    a. 0-25 books
    b. 26-100 books
    c. 101-200 books
    d. More than 200 books

APPENDIX B

Pre-Test

Rational numbers: Sum and subtraction

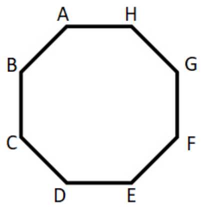
1. Consider the following numbers: 3/4; -5; 3; -2/5; -3.7; 0; -1.
   a. Write all the negative numbers: ___________________________.
   b. Calculate:
      i. -10+(-6)=
      ii. 14+13-(-5-2)=
2. Consider the following numbers: -1; 2.5; -1/2; 3(3/4); -4; 0
   a. Order the numbers from the largest to the smallest: ___________________________.
   b. Represent the numbers on the number line.
   c. Compute: -(-(-2))=
3.
   a. Are the following statements, correct? If they are not, correct them.
      i. $-1 \in \mathbf{Q}^{+}$
      ii. $\mathbf{Z} \subset \mathbf{Q}$
   b. Is the following representation, correct? If it is not, correct it.



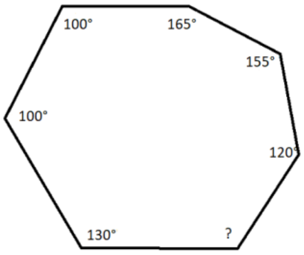
c. Write the following expression with symbols: add to the difference between 5.6 and -7 the sum between -2 and 3.

Polygons

1. Complete the statements.



   a. The polygon in figure is a ___________________________.
   b. The sides AB and AH are ___________________________ sides.
   c. The vertex F has ___________________________ non-adjacent sides.
2. Compute the unknown angle of the polygon in the following figure.



3. From a selected vertex of the regular polygon, we draw all the diagonals and thus divide the polygon into six triangles.
    a. How many vertices does this polygon have?
    b. What is the sum of the interior angles of this polygon?
    c. Find the measure of an interior angle of the polygon?

Rational numbers: Multiplication and division

1. Compute:
    a. (-8)×(-6)=
    b. 6×(-5)×(-4)=
    c. 5-3×2=
2. Compute:
    a. 2(3/4)(1(1/2)-2/5)=
    b. Solve the following equation: 4·x=-36
3. Compute:
    a. $\frac{1.5\times(-4)-(-0.8)}{-68}$=
    b. Solve the following equation: 2×|x|=6

Powers

1. Compute:
    a. 13^2=
    b. $(5/6)^2$=
    c. $\sqrt{81}$=
2. Compute:
    a. $\sqrt{36}+\sqrt{121}$=
    b. $(-3\cdot x\cdot y)^2$=
    c. a^3:a=
3. Compute:
    a. x^2=9
    b. $\sqrt{(3^2-7)\times 6+2^2}$=
4. Compute: $\frac{x^4.x^8.x}{x^7.x^6}$=

Relationship between variables

1. In the coordinate plane, plot the following points:

A (1, 2)

B (-3, 2)

C (0, -4)

2. Thomas paid for seven balloons 245 SIT. How much would he pay for 60 balloons?
3. Circle the letter that represents the table with the inverse proportion and complete the table.

(A)

x	6	10	14	8	
y	5	3	2.1		7

(B)

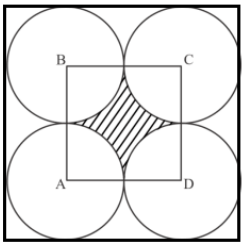
x	12		3	2	8
y	4	8	16		6

(C)

x	2	4		5	10
y	6	12	18		30

Circle and circumference

1. Determine the area of a circle with a radius equal to 2 dm.
2. In a circle with an area of 36π dm^2 we draw a central angle of 60°. Determine the length of the arc that is determined by the given central angle.
3. The square includes four circles, each with a ray of 5 cm. The centers of the circles are the vertices of a new square.



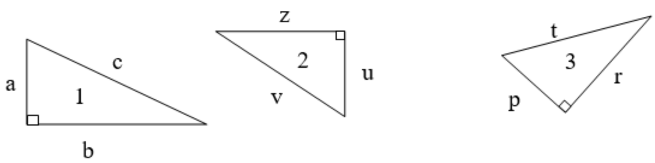
   a. Which is the length of the sides of the two squares?
   b. Which is the area of the shaded figure?

Computation with variables

1. Compute the following:
   a. x+x+x=
   b. $4y \cdot y^3 \cdot 3y^2$=
   c. 2a(a-b)=
2.
   a. Compute the following expression: 12x-13y-8x+15y=
   b. Compute the following expression for m=10 and n=-1:
   3(m-n)-2(m+n)=
3. Solve the following equation: (x-1) (x+2)-(x-1)·x=8

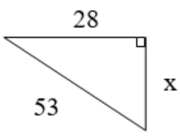
Post-Test

1. Draw a right triangle with legs 2.8 cm and 35 mm.
2. Complete the following table with respect to the figure:



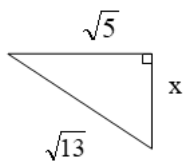
Triangle	Legs	Hypotenuse
1		
2		
3		

3. Determine the unknown side of the right triangle.

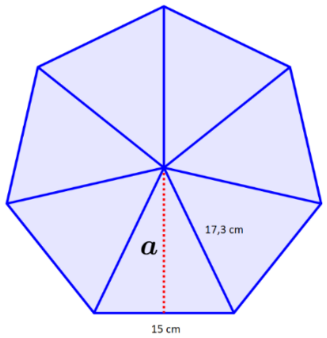


4. In a right triangle you know the length of the legs. Determine the perimeter and area of the triangle: s=40 cm; t=9 cm.

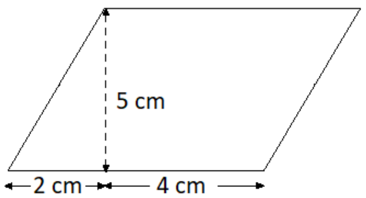
5. Is the triangle with sides a=8 cm, b=13 cm, and c=15 cm right-angled? (compute)
6. The area of a rectangle is 120 cm^2, and one of its sides is 8 cm. Determine the other side of the rectangle, the length of the diagonal, and the perimeter of the rectangle.
7. The perimeter of the triangle is 34 cm, and one of its sides is 12 cm. Determine the length of the diagonal and the area of the rectangle.
8. In a right triangle, the angle α=38°. The angle in vertex C is right. Which is the remaining angle?
9. Determine the perimeter, area, and length of the diagonal of a square with a side of 6 cm.
10. In the Cartesian plane draw the right triangle *AOB*, where *AOB*: *A*(0, -4), *O*(0, 0), *B*(-7, 0). Determine the length of all its sides. Then determine the perimeter and area of the triangle.
11. The length of the unknown side of the following right triangle is an irrational number. Determine it without the usage of a calculator.



12. Determine the area of the square whose diagonal is $\sqrt{2}$.
13. For which of the following polygons can you use the Pythagorean theorem? (Circle the correct answer)
    a. Regular hexagon
    b. Square
    c. Right triangle
    d. Rhombus
    e. Rectangle
14. Polona wants to create a scarf in the shape of a right isosceles triangle. One leg of the triangle is 6 dm. How much cloth will she need?
15. Determine the area of the heptagon in the figure.



16. Determine the perimeter and the area of the quadrilateral:



17. In a circle with a radius of 12.5 cm there is a rectangle, which length is 24 cm. Determine the height of the rectangle.
18. A ladder is spread out in the form of an isosceles triangle with legs of 41 dm, and it is spaced 1.8 m apart from the wall on the ground. How high can Andrej, who is 1.8 m tall and reaches 60 cm higher with his hands, reach with the ladder?

19. Determine the side b of a parallelogram with side a = 8 cm and the diagonal d = 10 cm.
20. How much m^2 do you need to create the following sail?

21. The beam has the shape of a rhombus with diagonals of 80 dm and 60 dm. We want to enclose it with concrete blocks. How many do we need if we have available ones that are 40 cm long?