PRIMARY SCHOOL STUDENTS´ STATISTICAL REASONING WHEN COMPARING GROUPS

Daniel Frischemeier
University of Paderborn, Germany
dafr@math.upb.de

Abstract: *This paper reports on the examination of group comparison strategies of primary school students in Germany (grade 4, age: 9-10 years), who were taught about group comparisons in a two-week statistics course. We present the design, the realization of the course and first empirical results of the evaluation. For the evaluation we gave the students a group comparison task before they attended the classroom activities and we gave them the same task after they had attended the classroom activities two weeks later. We collected the written notes from all students, analyzed the data with qualitative content analysis methods and compared the students´ outcomes and strategies before and after the participation in the course. The results show that students show more elaborated group comparison strategies after attending the course.*

INTRODUCTION

Concerned citizen need statistical skills to be able to participate in public decision processes. First steps of the development of these skills can be set in primary school and especially group comparisons can be a fruitful activity to enhance statistical reasoning at early stages (for definition of statistical reasoning, see Garfield & Ben-Zvi, 2008, p. 34). Since establishing the leading idea of data, frequency and chance in the German primary school curriculum (Hasemann & Mirwald, 2012), statistics has received increased attention. Recommended activities in this respect are posing statistical questions, collecting data, creating statistical displays like bar graphs, pie charts and dot plots, and reading and interpreting statistical displays like bar graphs, pie charts and dot plots. Comparing groups is an important and challenging activity in statistics which includes many fundamental statistical ideas like distribution, variation or representation and can build on the reading and interpreting of distributions of numerical variables. One common challenge for primary school students is to develop a global view on the distribution displayed in form of statistical graphs like stacked dot plots, because pupils often concentrate on local features like single points or extreme values (the maximum or the minimum) of a distribution (Bakker & Gravemeijer, 2004). Especially when dealing with stacked dot plots learners often focus on local features of the distribution and have problems to identify global features like center or spread. For instance modal clumps (see Konold et al., 2002 and Bakker, 2004) can serve as fruitful pre-stages for a concept of center (and also of spread) of a distribution and therefore can serve also as fruitful pre-concepts for group comparisons (comparing centers of distributions). Activities in this regard can already be taught in primary school and pre-comparison stages like modal clumps can help to make the transition from a local to a global perspective on distributions and may also help learners to compare groups even at early stages. The idea of the research study reported in this paper is to introduce primary school students to the modal clump concept, to lead them towards a global view of distribution and finally to provide them with a concept for comparing groups. In the following section we will shortly describe the design and the realization of lessons leading to group comparison activities in grade 4 (age: 9-10 years) in primary school. Then we will report on results of an accompanying empirical study which assesses the performance of students before and after attending to a teaching unit about comparing groups. The lessons were designed and taught by a preservice teacher for primary school within the scope of her Bachelor's Thesis (Breker, 2016).

A TEACHING UNIT TO ENHANCE STATISTICAL REASONING IN PRIMARY SCHOOL

In a Design-Based-Research setting (Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003), Breker (2016) designed, realized and evaluated a sequence of 13 lessons (45 minutes) for primary school students in grade 4 in Germany. Table 1 gives an overview of the content of each lesson. Before starting teaching, the pupils worked on a pretest where they were asked to compare groups

In: A Molnar (Ed.), *Teaching Statistics in a Data Rich World*
Proceedings of the Satellite conference of the International Association for Statistical Education (IASE),
July 2017, Rabat, Morocco. iase-web.org/Conference_Proceedings.php

in form of stacked dot plots (see Figure 1). Also at the end of the teaching unit the pupils worked on a posttest with the identical group comparison task of the pretest (see Figure 1). At the beginning of the teaching unit the pupils got to know about basics of data analysis (variable and values, lesson 1), collected data with a prepared survey in their school (lesson 2) and then got to know how to work with data cards and got to know several statistical displays like pie charts, bar graphs and stacked dot plots (lessons 3 & 4) in small datasets (grade 4 data).

No	Title of lesson	Content of lesson
1	Pretest We are introduced in data analysis	Pretest (see item in Figure 1) Pupils learn about variables (e.g., eye color) and values (e.g., blue, green, brown)
2	We collect our own data (class & school)	Data collection of the whole school with a prepared survey
3	We create our own displays (bar graphs) with class data	Separate, stack and order data cards, creating a bar chart with data cards
4	We create our own displays II (dot plots) with class data	Creating stacked dot plots with dots and finding a title for the display
5	We read our own displays with our class data	Reading the data and reading between the data on stacked dot plots, Pupils get to know modal clumps to identify characteristics of the distribution
6	We use TinkerPlots to create our own displays of our school data	Pupils adapt data card strategies to TinkerPlots to create bar graphs and stacked dot plots in larger datasets with TinkerPlots
7	We learn to compare groups I	Pupils are confronted with two distributions in form of stacked dot plots (equal sized) and get to know how to use modal clumps to compare two groups
8	We learn to compare groups II	Pupils are confronted with two distributions in form of stacked dot plots (non-equal sized) and get to know how to use modal clumps to compare two groups
9	We generate own statistical questions for comparing groups I	Pupils generate statistical questions leading to a group comparisons and conduct group comparisons with given data
10	We generate own statistical questions for comparing groups II	Pupils conduct group comparisons with given data and document their findings on posters.
11	We present our findings of our group comparisons I	Pupils present their findings to their classmates
12	We present our findings of our group comparisons II	Pupils present their findings to their classmates
13	Posttest	Posttest (see item in Figure 1).

Table 1: Overview of series of lessons.

In lesson 5 the pupils were taught to read the data, read between and beyond the data (see Friel, Curcio, & Bright, 2001) while being confronted with three types of statistical displays. Furthermore the pupils were guided from a local perspective to a global perspective in regard to the interpretation of distributions of a numerical variable in form of stacked dot plots using modal clumps. In lesson 6 the teacher used TinkerPlots (Konold & Miller, 2011) to produce these displays also for larger datasets (school datasets). In lessons 7-12 the pupils were taught about comparing distributions. The pupils learned how to identify modal clumps in single distributions of numerical variables, how to identify modal clumps in two distributions when comparing groups and how to compare the locations of the modal clumps. This was done in both group comparison settings (equal sized groups and non-equal sized groups). Finally, in lessons 9 & 10 the pupils were asked to work on own statistical investigations leading to group comparisons and then to conduct these group comparisons and to present and to discuss their findings in class in lessons 11 & 12. Further details about the design and the realization of the course can be read in Breker (2016).

RESEARCH QUESTIONS, TASK & METHOD OF STUDY

For this study the following main research question arises: *In which way are primary school students able to compare equal sized groups after attending to the teaching unit?* From this point one sub-research question emerges: *In which way does their performance for comparing groups improve after attending the teaching unit?*

To answer the research question and the sub-research question, we examine group comparison strategies of the primary school students before and after they were taught about group comparisons in a two weeks statistics course consisting of 13 lessons (see Table 1 for the overview). For the evaluation we gave the students a group comparison task before they have attended the classroom activities and we gave them the same task after they have attended the classroom activities two weeks later. We gave them a group comparison task where both groups are equal sized (see Figure 1) and also a group comparison task where both groups are non-equal sized. In this paper we will refer to the results on the group comparison task with the equal sized groups only. We can see this equal sized group comparison task in Figure 1.

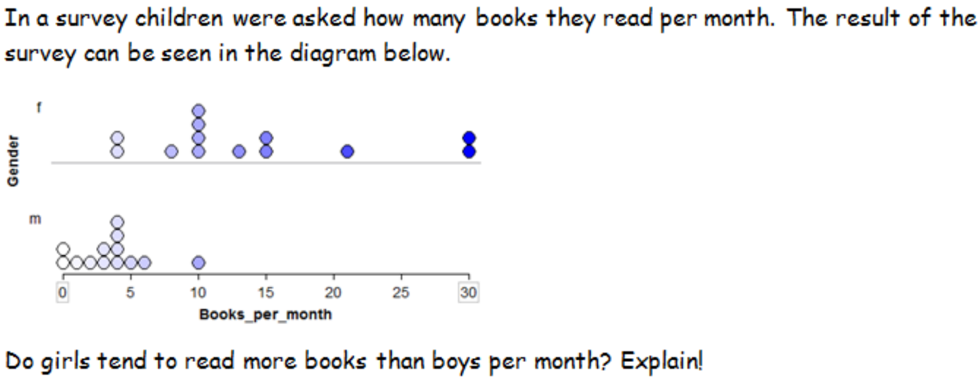

Figure 1: Group comparison task (equal size groups) from pre- and post-test.

There are several ways the pupils could solve this task. With their experiences from the teaching unit, they could identify modal clumps in both distributions and identify a shift between them to explain why girls tend to read more books than boys per month. Another possibility to find an adequate explanation is given by calculating and comparing the total score of both distributions and also by conducting p-based comparisons, which means that they choose a certain cut-point for both distributions and count and compare the number of data cases, which are larger/smaller than the cut-point. Note that the pupils were only used to additive but not to multiplicative reasoning.

DATA ANALYSIS & DISCUSSION

We collected the written notes from all students (n=11 before; n=11 after) and we have used qualitative content analysis methods (Mayring, 2015) to analyze the data, to derive different group comparison items, to rate the quality of these items and to compare the students´ outcomes and strategies before and after the participation in the teaching unit. For the coding procedure we used a mixed approach (Kuckartz, 2012) and generated the categories first from a deductive point of view and then – in a second step – refined them inductively. From the deductive point of view, we took into account the group comparison elements which are defined as sustainable group comparison items by Frischemeier (2017): Center, Spread, Skewness, Shift, p-based and q-based. We have left out q-based comparisons because they are too sophisticated for primary school pupils. With regard to comparisons of center, we also added "total score" of two groups as comparison element, which is a sustainable element when the two groups are equal-sized. From an inductive point of view we added idiosyncratic and extreme values like maximum/minimum or outliers of a distribution. We also distinguished shift comparisons, where the shift of multiple data points (not mentioned as modal clump) of both distributions is compared (Shift Ps) and where the shift of modal clumps of both distributions is compared (Shift MC). In a final third step we have rated "no explanation" and "idiosyncratic explanation" with (-) and explanations based on extreme values with (o) because extreme values are no global features of distributions. The other explanations (p-based, Skewness, Spread, Shift Ps, Shift MC, Total Score and Center) are all rated as (+) since these elements take into account global features of distributions and can be seen as adequate group comparison elements when comparing two equal sized groups. The overview of group comparison items can be seen in Table 2.

Group comparison Item (Rating)	Definition	Example
None (-)	No item for comparison is used.	-
Idiosyncratic (-)	The item used for comparison is idiosyncratic.	“The boys read more, because you see it here.” (Lorenz)
Extreme values (o)	Extreme values like maximum/minimum or outliers of a distribution of a numerical variable are used to explain the difference between the two groups.	“The girls read more books, because there are 2 girls who read 30 books per month.” (Peter)
p-based (+)	Learners choose a cut-point in both distributions and count and compare the number of data cases larger/smaller than the cut-point in both distributions	“Girls read more books per month, because there are more points on the larger values.” (Marc)
Skewness (+)	The difference between the two groups is explained with differences in skewness.	No example in data
Spread (+)	The difference between the two groups is explained with differences in spread.	“Because the points of the girls are not so in one place compared to the points of the boys. The points of the girls are more distributed towards the higher values.” (Sarah)
Shift Ps (+)	Learners compare the location of more than one point in both distributions.	“Girls read more because the points are more right.” (Noel)
Shift MC (+)	Learners identify modal clumps in both distributions and compare the shift between both modal clumps.	“Girls [read more] because the clump is more right.“ (Marc)
Total Score (+)	Learners calculate the total score of each distribution and compare both sums.	No example in data
Center (+)	The difference between the two groups is explained with differences in center (mean/median).	No example in data

Table 2: Overview of group comparison items with definitions and examples.

RESULTS

In Table 3 we compare students’ performance on the group comparison task before and after the students have attended the teaching unit. At first we can say that even before participating in the teaching unit, many pupils showed a correct understanding about “whether girls tend to read more books per month than boys”, because 9 of 11 students were able to solve the group comparison task successfully. But only 2 of the 9 students, who gave the correct answer, were able to give a good explanation (+) like p-based or spread. Nobody used shift features to decide whether girls tend to read more books than boys in the pretest. In addition 4 of the 9 students explained the difference with an explanation rated medium (o), concentrating on local features like the maximum of the distribution. Finally 3 of the 9 students gave no or a wrong explanation (-). In regard to the results of the posttest we can say, that in the posttest 10 of 11 students gave a correct answer to the question whether girls tend to read more books than boys. The real improvement can be seen in the last column (Explanation after) in Table 3 compared to the column (Explanation before), because 8 of the 10 students who gave a correct answer used adequate explanations (Shift Ps and Shift MC) in the posttest to decide that girls tend to read more books per month than boys.

No	Name	Correct (before)	Explanation (before)	Correct (after)	Explanation (after)
1	Peter	Yes	Maximum (o)	Yes	Shift Ps (+)
2	David	Yes	None (-)	Yes	Idiosyncratic (-)
3	Lorenz	No	Idiosyncratic (-)	Yes	Shift Ps (+)
4	Lars	Yes	Maximum (o)	Yes	Shift MC (+)
5	Marc	Yes	p-based (+)	Yes	Shift Ps (+)
6	Noel	Yes	None (-)	Yes	Shift Ps (+)
7	Sarah	Yes	Spread (+)	Yes	Shift Ps (+)
8	Fiona	Yes	Maximum (o)	Yes	Shift MC (+)
9	Christian	Yes	None (-)	Yes	None (-)
10	Johannes	No	Spread (-)	No	Idiosyncratic (-)
11	Josefine	Yes	Maximum (o)	Yes	Shift MC (+)

Table 3: Overview of group comparison performance (and change: before/after) of all pupils.

We can say that there is a slight improvement of the correctness of answers (9 of 11 correct in pretest compared to 10 of 11 correct in the posttest) and a huge improvement in the quality of the explanations. As an example for the improvement of explanations let us have a look on the written note of Lars. In the pretest Lars solved the group comparison task correctly ("girls read more books per month") and gave an explanation for his statement (Figure 3) in regard to extreme values ("Girls read more books per month because at 30 books, there is a "0" for the boys and a "2" for the girls"). Lars compared the distributions taking into account local features like single points of the distribution (30 books per month). Then he compared the number of girls reading 30 books per month (=2) with the number of boys reading 30 books per month (=0).

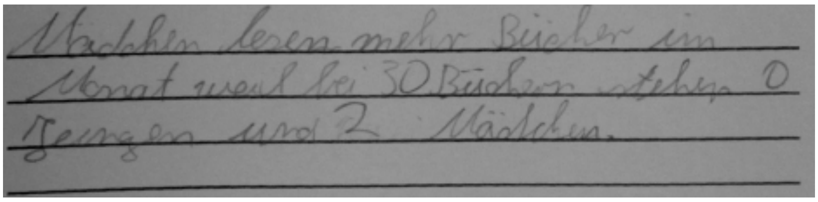

Figure 3: Written note of Lars (pretest).

In the posttest, Lars also solved the group comparison task correctly and gave an adequate explanation for his statement ("Girls because the clump is more right") as can be seen on the left side of Figure 4.

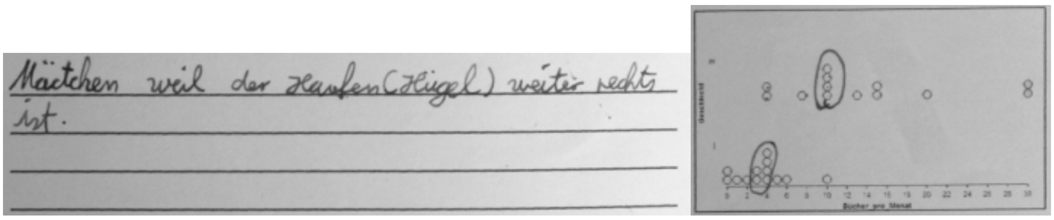

Figure 4: Written note of Lars (posttest) and modal clumps in the distributions (right).

Having a look at Lars' graph (Figure 4 right), we also see that he marked two areas to identify modal clumps. In the distribution of girls (upper distribution in Figure 4 right) he marked the data on 10 books with a curl and also marked the data points which are smaller than 10 books per month with dots. Lars also marked the data of 3 and 4 books per month in the distribution of boys (lower distribution in Figure 4 right) and compared the location of both modal clumps to explain that girls tend to read more books than boys per month.

CONCLUSIONS & RECOMMENDATIONS

The study reported in this paper offers interesting insights into fourth graders´ group comparison strategies for the further development of group comparison activities. We can conclude that it is possible to establish first notations of comparing groups and that group comparison activities can also be implemented in primary school classroom (after introducing basics of data analysis, stacked dot plots, etc.). Even in the pretest, most of the pupils have shown that they were able to decide whether girls tend to read more books than boys, but the explanations of many of the pupils have been wrong or focusing on local features. After attending to the teaching unit the quality of explanations has improved massively. This can be seen as a first step to enhance sophisticated statistical reasoning in primary school. In regard to task design it could be meaningful to use larger samples and therefore have a better possibility to define modal clumps. It was remarkable that in the posttest many pupils used modal clumps to identify and compare the centers of the distribution – although the pupils who applied adequate other explanations like p-based comparisons (Marc) or spread (Sarah) in the pretest did not build on their good reasoning and used modal clumps for explanation instead. So in regard to the teaching unit it would be positive to make more interventions than modal clumps for identifying and comparing the centers of the distributions, so that pupils use other adequate explanations like spread or p-based to compare groups and that pupils (like Marc or Sarah) still use these kinds of explanations and improve them instead of moving to another explanations like Shift Ps/Shift MC. Further research has to deal with students reasoning in group comparison settings which are non-equal sized.

REFERENCES

Bakker, A. (2004). *Design research in statistics education - On symbolizing and computer tools.* (Unpublished Doctoral Dissertation). University of Utrecht, Utrecht, the Netherlands.

Bakker, A., & Gravemeijer, K. (2004). Learning to reason about distributions. In D. Ben-Zvi & J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 147–168). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Breker, R. (2016). *Design, Durchführung und Evaluation einer Unterrichtseinheit zur Entwicklung der Kompetenz „Verteilungen zu vergleichen“ in einer 4. Klasse unter Verwendung der Software TinkerPlots und neuer Medien.* (Unpublished Bachelor of Education thesis). University of Paderborn, Paderborn, Germany.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design Experiments in Educational Research. *Educational Researcher, 32*(1), 9–13.

Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making Sense of Graphs: Critical Factors Influencing Comprehension and Instructional Implications. *Journal for Research in Mathematics Education, 32*(2), 124–158.

Frischemeier, D. (2017). *Statistisch denken und forschen lernen mit der Software TinkerPlots*. Wiesbaden, Germany: Springer Fachmedien Wiesbaden.

Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning. Connecting Research and Teaching Practice*. Dordrecht, The Netherlands: Springer.

Hasemann, K., & Mirwald, E. (2012). Daten, Häufigkeit und Wahrscheinlichkeit. In G. Walther, M. van den Heuvel-Panhuizen, D. Granzer, & O. Köller (Eds.), *Bildungsstandards für die Grundschule: Mathematik konkret* (pp. 141–161). Berlin, Germany: Cornelsen Scriptor.

Konold, C., & Miller, C. (2011). TinkerPlots 2.0 [Computer software]. Emeryville, CA: Key Curriculum Press.

Konold, C., Robinson, A., Khalil, K., Pollatsek, A., Well, A., Wing, R., & Mayr, S. (2002). *Students' use of modal clumps to summarize data.* Paper presented at the Sixth International Conference on Teaching Statistics, Cape Town, South Africa.

Kuckartz, U. (2012). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstüzung.* Weinheim, Basel, Germany: Beltz Juventa.

Mayring, P. (2015). Qualitative content analysis: theoretical background and procedures. In A. Bikner-Ahsbahs, C. Knipping, & N. Presmeg (Eds.), *Approaches to qualitative research in mathematics education* (pp. 365-380). Dordrecht, The Netherlands: Springer.