ENHANCING DATA LITERACY AT AN EARLY STAGE: PRIMARY SCHOOL PUPILS´ PRECURSOR IDEAS OF COVARIATION

Martin Andre[1], Daniel Frischemeier[2], & Christoph Schneider[1]
[1]University College of Teacher Education Tyrol, Austria
[2]University of Münster, Germany
martin.andre@ph-tirol.ac.at

Covariational reasoning can already be enhanced in primary school. Anyway, given that pupils do not know conventional notations like coordinate systems in the first grades, addressing ideas of covariation is a challenging task. In our study, we show that 1st graders' (6-7 years) ideas of covariation already exist in a precursor way and can be addressed by specific tasks based on data from Gapminder's photo database Dollar Street. Using a design-based research approach, we developed and implemented semi-digital learning trajectories to enhance pupils' data literacy at a very early stage. Interactions between young learners are observed to gain insights (a) into their covariational thinking when working on these specific tasks and (b) for iteratively re-designing the learning trajectories. Our observations define an early starting point for learning processes of covariational reasoning and the learning trajectories give opportunities to address issues of sustainable development.

INTRODUCTION

Many decisions in politics, social sciences, medicine, and economics are based on data. Especially in these days of the Covid-19 pandemic, a competent use of data is becoming increasingly important. Citizens' understanding of data is essential to ensure a joint, purposeful approach to face the pandemic. In general, data literacy is a crucial ingredient to become an engaged citizen (Engel, 2017). Therefore, statistical thinking should be encouraged as early as possible, preferably in primary school (Leavy et al., 2018). For adequate understanding and motivating enquiry, young pupils can already explore meaningful and large data sets at primary level (Frischemeier, 2020). To explore meaningful and large data sets and to find interesting insights and relations in the data, the use of digital data analysis tools is useful and necessary. The ProCivicStat project has developed plenty materials (https://iase-web.org/islp/pcs/) to teach and bring fundamental statistical ideas in the frame of civic statistics into classrooms, which comprise the field of migration, health, and poverty (Engel, Gal, & Ridgway, 2016). Variation and covariation are fundamental ideas in statistics (Burrill & Biehler, 2011) and these covariational issues are – in their formal way – part of the middle or high school curriculum. In this paper we want to share our research and experiences on how to develop first precursor ideas of covariation in the context of the sustainable development goals with first grade pupils (6-7 years).

THEORETICAL BACKGROUND AND LITERATURE REVIEW

Our aim to develop first precursor ideas of covariation at an early stage is ambitious. We lay our theoretical background on three different dimensions. The first dimension is the statistical content (covariation), the second dimension is the context (sustainable development), and the third dimension is the digital tool used in the exploration phase (Gapminder). We will elaborate on all three dimensions in the following shortly.

Statistical dimension

According to Burrill and Biehler (2011) fundamental statistical ideas are data, distribution, representation, variability, and others. To educate young learners who do not possess formal statistical knowledge, Konold (2002) strongly suggests building on their informal ideas as conceptual foundations for introducing formal procedures later in school. With regard to associations in data, Konold et al. (1997) distinguish three types: categorical vs. categorical, numerical vs. categorical, and numerical vs. numerical. We did not find any studies which address students' covariational reasoning using qualitative data such as photos vs. quantitative data until now. In our study we will introduce this case as well as we will focus on the association between two numerical variables.

Moritz (2004) found four levels of verbal and numerical graph interpretation regarding learners covariational reasoning, increasing from level 0 – 'nonstatistical' and level 1 – 'single aspect' over level 2 – 'inadequate covariation' to level 3 – 'appropriate covariation'. Focusing on level 3, Aridor and

In: R Helenius, E Falck (Eds.), *Statistics Education in the Era of Data Science*
Proceedings of the Satellite conference of the International Association for Statistical Education (IASE),
Aug-Sept 2021, Online conference. iase-web.org/Conference_Proceedings.php

Ben-Zvi (2019) identified two major aspects when students developed their aggregate reasoning with covariation (ARwC), i.e., reasoning with modeling and reasoning with variability. Moreover, the authors found the growing samples pedagogy to be supportive for students' development of ARwC.

The standard display for visualizing covariation in the sense of focusing on two numerical variables are scatterplots – but many learners face difficulties to interpret scatterplots (Biehler et al. 2018) – and have problems to go beyond a mere reading the data (e.g., Friel et al. 2001; Moritz, 2004). Therefore Konold (2002) for instance suggests using alternatives to scatterplots as a pre-stage – preferably to refer the interpretation of scatterplots to well-known situations –which are easier to interpret, e.g., group comparisons. A similar approach can be found in Cobb et al. (2003). Following this direction, our idea is to facilitate the interpretation of covariation and refer to already known situations as well as to refer relationships in quantitative data to qualitative situations.

With this study we want to explore options for addressing the idea of covariation in a precursor stage, combining qualitative data such as photos of living standards and quantitative data, in our case income data, in a first step. Building upon that, in a second step, we want to observe whether pupils intuitively describe ideas of covariation when examining a scatterplot of the same context.

Technological dimension

The use of software is one fundamental aspect for setting up adequate statistical reasoning learning environments (Garfield & Ben-Zvi, 2008). Software and digital data analysis tools enable learners to explore large and meaningful datasets and they reduce the cognitive load so that young learners can concentrate on statistical reasoning rather than on technical procedures. In addition, these tools help learners to generate multiple visualizations of data. Data visualization tools like *Gapminder* (Rosling et al., 2005) are freely available and as Biehler et al. (2018) mention the "use of technology may also enhance covariational thinking" (p. 182). We see especially Gapminder (www.gapminder.org) as adequate tool for this purpose because Gapminder includes relevant civic statistics data and offers the possibility to explore relationships between multiple variables. Moreover, Gapminder's photo database *Dollar Street* (Rosling, 2015) opens opportunities to explore qualitative aspects of data related to income situations. "[…] It seemed natural to use photos as data so people can see for themselves what life looks like on different income levels. Dollar Street lets you visit many, many homes all over the world" (Rosling, 2015). As these Gapminder's data visualizations are strongly connected to income levels and, moreover, young learners already have a conceptual understanding of the terms 'rich' and 'poor', in our study we focus on data visualizations including the variable 'income'.

Context dimension

In 2015, 193 member states of the United Nations enacted the 17 *sustainable development goals* (*SDGs*) as part of the *Agenda 2030* (United Nations, 2015), which should mediate between social, ecological, and economic issues, worldwide. With regard to sustainable development and education for sustainable development (ESD), context and contextual knowledge about the environment and the economy are crucial parts of developing competences in ESD (Rieckmann, 2018). In their meta-study, Taylor et al. (2019) point out that teachers encounter major problems integrating ESD into their lessons. The aim of ESD is to support learners in developing competences towards participating in sustainable development by raising awareness of problems associated with the topic as well as critically reflecting these issues (Rieckmann, 2018) considering individual points of views and global perspectives. Based upon these aims, Rieckmann (2018) defined several key competences for ESD, e.g., the competence to analyze complex systems or forward-looking thinking. In some parts these competences are in accordance with basic statistical ideas such as, awareness of the nature of data, comparing groups by centre and spread, ideas of statistical inference or the idea of building statistical models as to regressions.

Researchers agree that a context meaningful to learners is a crucial concern for teaching statistics (e.g., Aridor & Ben-Zvi, 2019). According to Pfannkuch and Rubick (2002), context and contextual knowledge is "essential for conjecturing possible relationships within the set of data. Both contextual and statistical knowledge influenced students' understanding and interpretation of the data" (p. 16). In this regard Langrall et al. (2011) showed in their study with middle school students the positive effects of contextual expertise on in-depth data exploration, especially on drawing and justifying conclusions. Addressing issues of sustainable development, Andre et al. (2020) also found the context and contextual knowledge to be essential for pupils' intuitive approaches to statistical ideas of covariation. Therefore, in our current study we focus again on issues of sustainable development as the context of pupils' data explorations, addressing single SDGs.

Taking into account the three outlined dimensions, we can summarize that we will concentrate on the statistical topic covariation, use the digital tool Gapminder and Gapminder's Dollar Street and the context of the relationship between income and some other issues of sustainability. Building on that, we may generate the following research question: In how far is it possible to develop covariational thinking at an early stage (6 ages) using Gapminder's data visualizations and the photo database Dollar Street while addressing issues of sustainable development?

METHODS AND IMPLEMENTATION

This report deals with the first phase of a three years' design-research project (Bakker, 2018) aiming to combine education for sustainable development and statistics education in primary school. Based on earlier studies (Andre et al., 2020), our aim in the current project is to develop learning trajectories for primary school pupils to enhance their statistical reasoning while addressing various sustainable development goals in depth. From our earlier studies we already know the ideas of correlation and regression to be intuitively evident in many of the Gapminder bubble charts. Therefore, besides designing practicable learning trajectories, we target to further develop theories on young learners' covariational thinking with our project. Primary school teacher students were involved in the design process and the classroom implementation of the learning trajectories.

As part of a seminar on statistics and probability, six primary school teacher students were guided to create virtual learning trajectories addressing young learners' statistical thinking and especially their covariational thinking. Due to COVID-19 regulations in schools, the learning trajectories were elaborated as homepages so that the implementation could take place both, in presence and virtually. In advance, students were introduced to the theoretical background of young learners' statistics education and the software h5p. Using this software, students in pairs of two created a digital learning trajectory, addressing several topics of sustainable development, on the one hand, and statistical as well as data issues on the other hand. Especially focusing on sustainable development goals, specifically goal six – clean water and sanitation, goal twelve – responsible consumption, and goal fifteen – life on land were emphasized.

In this report, we focus on a three weeks' classroom implementation with four teacher students in May 2021 at an Austrian Elementary School in three working blocks of 90 minutes' duration each. The class was attended by a total of 18 children aged 6 to 7 years, in the gender split of 12 boys and 6 girls. During the implementation phase, teacher students worked in pairs with a group of 9 children each on their designed learning trajectory. These two groups were further divided into small groups of three children working together on the learning trajectory with one digital device. Students' and pupils' statements were analyzed and translated into English language afterwards.

In the first phase of the implementation, pupils were introduced to the topic, and they got familiar with the virtual learning trajectories. Self-explanatory videos and posters were used to provide children with information on each topic. Moreover, bar charts were introduced to the pupils, and they created and discussed their own bar charts on given data, in this first phase. In the second phase, the learning trajectories differed from each other. While the focus of the first group was on the display of contextual data in a circle diagram, the other group focused on the idea of distribution when exploring diverse issues in the photo database Dollar Street and relating these issues to their own lives. Thus, children graphically allocated their personal facilities like toys, cutlery, or toilets, in the Dollar Street. The contents discussed by each group were picked up and deepened in the virtual learning trajectory using task formats such as cloze texts, multiple choice questions and concrete question-answer options (true/false). In our research, we did not deepen on these tasks concerning statistical content that is not directly connected to covariational reasoning. Anyway, these tasks were included to strengthen pupils' confidence with the context of the addressed SDGs and, moreover, they got to know various data visualizations concerning these issues.

As a central task for our study, all pupils' groups should assign different pictures of the Dollar Street showing, for example, bathrooms or living rooms to the monthly income which could be categorized in low, middle, or high income. Moreover, bubble charts on the specific SDGs-topic were shown to the children, displaying the relation of income and CO_2-emissions or access to sanitation. The teacher students accompanied these two tasks carefully, discussing various aspects with the pupils and guiding them with questions. Working on these tasks together with the teacher students, some pupils' statements were recorded to be analyzed later. In the third and last phase, pupils worked in groups to finalize the tasks and to reflect their working process together with the teacher students.

In our report, we focus on the second phase and especially on the two tasks shown in figure 1. The first task shown on the left-hand side was to allocate the photos of a bathroom to the according income. As a second task, the graphic on the right-hand side, was introduced to the children and they were interviewed by the teacher students regarding their perception of the relationship between the average income in a country, which is shown on the x-axis, and the share of people who have access to basic sanitation at least.

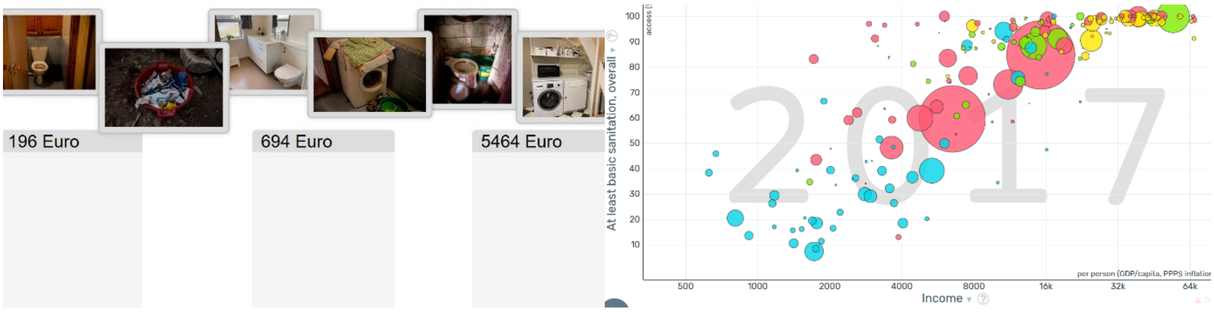

Figure 1. Left: Example of a task relating qualitative data (toilets, bathrooms, and laundry) and quantitative data (income) to each other. Right: Gapminder's bubble chart in the same context (income and access to basic sanitation) to be discussed by the pupils.

Performing a qualitative content analysis (Kuckartz, 2012), we examined the collected data in two directions. On the one hand, we evaluated the children's actions and statements in terms of their achievement of the previously stated learning objectives for each task, building two categories, i.e., achieved and missed. On the other hand, we analyzed the interviews with the children building the categories 'covariation', 'causality and variance', and 'misconceptions and difficulties'.

FIRST RESULTS AND DISCUSSION

Working on a drag-and-drop task, the pupils had to connect three income levels with corresponding sanitary facilities (see figure 1). During our observations, all children managed to find the correct relationships, arguing for example: "Because, they have certainly got a lot of money a month for that nice toilet" or "The toilet and the bathroom is really large." Another pupil justified the choice of a small income: "Because they get almost no money [...] it's just something where you have to sit outside."

Some pupils also mentioned aspects of causality and variance, stating that richer people do not necessarily have to own a nice bathroom as well, as this dialogue shows:
Interviewer: "Do you think that families with more money always have a nicer bathroom?"
Child: "No!"
Interviewer: "Don't you think so? So it could be that people with less money have a nice toilet?"
Child: "It could be. (...) Some people don't buy a nicer toilet but spend more money as family budget."
Interviewer: "So don't you think that high income always comes together with a nice toilet?"
Child: "No!"

Another pupil answered the question „The richer a person is, the more likely does the person have an own bathroom?" with "Yes, but just more likely". The question "Do rich people always have nice toilets?" was answered by another pupil with "Could be, but not necessarily". Thus, an important aspect of our analysis concerned causality and variance, which was inherent in many statements.

In a later task, pupils were interviewed on a bubble chart. One group of teacher students used a bubble chart in the context sanitation facilities (see fig. 1). Another students' group interviewed the pupils on a Gapminder's bubble chart showing the covariation of income and CO_2-emission. With the help of the students, in both cases many children were able to read the two dimensions of the scatterplot graphics, correctly interpreting the position of certain points in the scatterplot. We also found some pupils' statements like "When they earn a lot of money, they can buy a toilet and when they don't have much money, they can't." Examining the scatterplot, another pupil reasoned "Because through the money they buy toilets and if they have so much money, then additionally [they buy a] tidy apartment." Dealing with the covariation between CO_2-emissions and income of the various countries, we found, too, that the interviewed children correctly interpreted the position of points in the scatterplot but in this group, pupils did not formulate sentences of the form "the higher … the more". Asking our teacher students for possible reasons, they supposed that CO_2-emissions were an unknown term to the children and therefore the children could not find appropriate argumentations or reasons for such a covariation.

Additionally, we found that children were also confused by the large amount of information given in the Gapminder's bubble chart in the beginning. Many children interpreted the size of the bubbles, which actually stands for the population size, as a country's income level. Others were attracted by the colors of the bubbles and examined more geographic issues. Only with assistance of the teacher students, some of the children could also interpret most information of the bubble chart appropriately.

CONCLUSION

Aiming to explore opportunities to address young children's precursor ideas of covariation, with our study, we introduced the idea of integrating qualitative data in children's learning processes of covariation. Thus, all pupils were able to reason with covariation when examining income and photos of people's living standards such as bathrooms, toilets, or laundries. Without deep scaffolding, they argued relations between income levels and sanitation facilities, but they were carefully when expressing causalities. Applying the classification by Moritz (2004) on the children's precursor ideas of covariation, many of the children reached level 2 or 3, interpreting the covariation between photos of living standards and income appropriately in many cases. Furthermore, as described by Aridor & Ben-Zvi (2019), we also found reasoning with variability to be important for their development of covariational reasoning.

Working with Gapminder's bubble charts in a second step, children needed guidance and scaffolding in many places, sometimes they misinterpreted aspects of the graphics, and especially in the case of CO_2-emissions they did not clearly express reasoning with covariation, although the scatterplot is showing better the idea of covariation than the one displaying income and sanitation. The children did not fully understand what CO_2-emissions are, while the context of sanitation was known from the tasks before. Therefore, we conclude that the level of children's abilities to interpret scatterplots strongly depends on children's knowledge of and confidence with the context. Anyway, with no statistical focus or a focus on single points, their abilities to interpret scatterplots did not exceed level 1 in many cases.

From our observations we conclude that the approach of including qualitative data to address young children's precursor ideas of covariation can contribute to their development of covariational reasoning. Based on these experiences and to deepen insights into the children's reasoning processes, a re-design of the learning trajectories will include negative covariation (e.g., Moritz, 2004) and aspects of growing samples pedagogy (Aridor & Ben-Zvi, 2019). Furthermore, we will deepen data collection regarding pupils' reasoning with modeling and variability. Finally, we will continue to address the context of sustainability, giving opportunities to deepen teacher students' and pupils' awareness and knowledge of these issues and opening a broad field of topics highly relevant for future generations.

ACKNOWLEDGMENTS

We want to thank our teacher students for their contributions in generating and implementing the learning trajectories. Moreover, we are very grateful for the detailed and constructive reviews, which considerably helped to improve our paper.

REFERENCES

Andre, M., Lavicza, Z., & Prodromou, T. (2020). Integrating 'education for sustainable development' in statistics classes: Visual analysis of social and economic data with gapminder. In P. Arnold (Ed.), *New Skills in the Changing World of Statistics Education. Proceedings of the Roundtable conference of the International Association for Statistical Education (IASE).* International Association for Statistical Education.

Aridor, K., & Ben-Zvi, D. (2019). Students' Aggregate Reasoning with Covariation. In G. Burrill & D. Ben-Zvi (Eds.), *Topics and Trends in Current Statistics Education Research* (pp. 71–94). Springer International Publishing. https://doi.org/10.1007/978-3-030-03472-6_4

Bakker, A. (2018). *Design research in education: A practical guide for early career researchers*. Routledge, Taylor & Francis Group.

Biehler, R., Frischemeier, D., Reading, C., & Shaughnessy, J. M. (2018). Reasoning About Data. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International Handbook of Research in Statistics Education* (pp. 139–192). Springer. https://doi.org/10.1007/978-3-319-66195-7_5

Burrill, G., & Biehler, R. (2011). Fundamental Statistical Ideas in the School Curriculum and in Training Teachers. In C. Batanero, G. Burrill & C. Reading (Eds.), *Teaching Statistics in School Mathematics – Challenges for Teaching and Teacher Education* (pp. 57-69). Springer. https://doi.org/10.1007/978-94-007-1131-0_10

Cobb, P., McClain, K., & Gravemeijer, K. (2003). Learning About Statistical Covariation. *Cognition and Instruction*, *21*(1), 1–78. https://doi.org/10.1207/S1532690XCI2101_1

Engel, J., Gal, I., & Ridgway, J. (2016). Mathematical Literacy and Citizen Engagement: The Role of Civic Statistics. Paper presented at the 13th International Congress on Mathematical Education, Hamburg.

Engel, J. (2017). Statistical Literacy for active Citizenship: A Call for Data Science Education. *Statistics Education Research Journal*, *16*(1), 44-49.

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. https://doi.org/10.2307/749671

Frischemeier, D. (2020). Building Statisticians at an Early Age - Statistical Projects Exploring Meaningful Data in Primary School. *Statistics Education Research Journal 19*(1), 39-56.

Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and practice*. Springer. https://doi.org/10.1007/978-1-4020-8383-9

Konold, C., Pollatsek, A., Well, A., & Gagnon, A. (1997). Students Analyzing Data: Research of Critical Barriers. In J. B. Garfield & G. Burrill (Eds.), *Research on the Role of Technology in Teaching and Learning Statistics: 1996 Proceedings of the 1996 IASE Round Table Conference* (pp.169-190.). International Statistical Institute.

Konold, C. (2002). Alternatives to scatterplots. *Proceedings of the Sixth International Conference on Teaching Statistics*. Cape Town, South Africa.

Kuckartz, U. (2012). *Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung.* Beltz-Juventa.

Langrall, C., Nisbet, S., Mooney, E., & Jansem, S. (2011). The role of context expertise when comparing data. *Mathematical Thinking and learning, 13*, 47-67. https://doi.org/10.1080/10986065.2011.538620

Leavy, A., Meletiou-Mavrotheris,M., & Paparistodemou, E. (2018). *Statistics in early childhood and primary education: Supporting early statistical and probabilistic thinking*. Springer. https://doi.org/10.1007/978-981-13-1044-7

Moritz, J. B. (2004). Reasoning about covariation. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 227–256). Kluwer Academic Publishers.

Pfannkuch, M., & Rubick, A. (2002). An exploration of students' statistical thinking with given data. *Statistics Education Research Journal*, *1*(2), 4–21.

Rieckmann, M. (2018). Die Bedeutung von Bildung für nachhaltige Entwicklung für das Erreichen der Sustainable Development Goals (SDGs). *Zeitschrift für internationale Bildungsforschung und Entwicklungspädagogik*, (2), 4–10. http://dx.doi.org/10.31244/zep.2018.02.02

Rosling, A.R. (2015). *About Dollar Street*. https://www.gapminder.org/dollar-street/about

Rosling, H., Ronnlund, A.R., & Rosling, O. (2005). New software brings statistics beyond the eye. In E. Giovannini (Ed.), *Statistics, knowledge and policy: Key indicators to inform decision making* (pp.522–530). Organization for Economic Co-Operation and Development.

Taylor, N., Quinn, F., Jenkins, K., Miller-Brown, H., Rizk, N., Prodromou, T., Serow, P., & Taylor, S. (2019). Education for sustainability in the secondary sector - a review. *Journal of Education for Sustainable Development, 13*(1), 102–122. https://doi.org/10.1177/0973408219846675

United Nations (2015). *Sustainable development goals*. Retrieved from https://sustainabledevelopment.un.org