



HAL
open science

Stepwise development of statistical literacy and thinking in a statistics course for elementary preservice teachers

Daniel Frischemeier, Rolf Biehler

► To cite this version:

Daniel Frischemeier, Rolf Biehler. Stepwise development of statistical literacy and thinking in a statistics course for elementary preservice teachers. CERME 10, Feb 2017, Dublin, Ireland. hal-01927856

HAL Id: hal-01927856

<https://hal.science/hal-01927856v1>

Submitted on 20 Nov 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Stepwise development of statistical literacy and thinking in a statistics course for elementary preservice teachers

Daniel Frischemeier & Rolf Biehler

Paderborn University, Institute of Mathematics, Germany; dafr@math.upb.de

In this paper we describe the design, realization and evaluation of a course for elementary preservice teachers, applying the PPDAC-cycle (Wild & Pfannkuch, 1999), using innovative methods and digital tools like TinkerPlots (Konold & Miller, 2011). We will refer to design principles of the course and show in which way a stepwise development of statistical literacy and thinking with TinkerPlots works in cooperative learning environments.

Keywords: Elementary preservice teachers, design based research, cooperative learning, statistical literacy and thinking, TinkerPlots.

Introduction

Since the implementation of the leading idea “Data, Frequency and Chance” (Hasemann & Mirwald, 2012) in mathematics classrooms in primary schools in Germany, statistics has become a central topic in primary school. This has set requirements not only for schools and teachers, but also for universities who have to educate preservice teachers in statistics for their upcoming school career. Requirements for teacher education in statistics can be found on German national level (e.g. AK Stochastik, 2012), and on international level (e.g. Batanero, Burrill & Reading, 2011). Two important aspects appear at both levels: applying a whole data analysis cycle (like PPDAC, see Wild & Pfannkuch, 1999) and analyzing data with digital tools (see Biehler, Ben-Zvi, Bakker & Makar, 2013). We decided to use TinkerPlots (Konold & Miller, 2011) for our purposes, since it is easy to learn, no formulas are needed, and it enables learners to create multiple representations of data. In our sense, TinkerPlots can serve as educational software for pupils from grade 4, as software for teachers for analyzing data, and as medium for demonstration purposes in classroom. This was our motivation to design, realize and evaluate a statistics course for elementary preservice teachers with TinkerPlots on the basis of the Design Based Research paradigm (Cobb, Confrey, diSessa, Lehrer & Schauble, 2003).

Our course to develop statistical literacy and thinking with TinkerPlots

The main goal of this course is to develop statistical literacy and thinking components (for a definition see Garfield & Ben-Zvi, 2008, pp. 34) and the technological knowledge of our participants. At Paderborn University in Germany elementary preservice teachers for mathematics attend an obligatory course “Elementary Statistics”, which is about data analysis, combinatorics and probability theory. Due to limits of time, there is no space for going through a whole data analysis cycle like PPDAC or to do further data explorations in multivariate datasets. For that reason our course was designed taking into account the principles of the “Statistical Reasoning Learning Environment” (Garfield & Ben-Zvi, 2008), to expand preservice teachers’ knowledge in data analysis and to introduce a new tool to them, which better fits to primary and lower secondary school. Fundamental ideas realized in our new designed course are to “focus on developing central statistical ideas”, to “use real and motivating data sets”, to “use classroom activities to support the development of students’ reasoning”, the integration of “appropriate technological tools”, to “promote classroom discourse that includes statistical arguments and sustained exchanges that focus on

significant statistical ideas” and the “use of formative assessment” (see Garfield & Ben-Zvi, 2008, p. 48). On the paradigm of the PPDAC-Cycle (Wild & Pfannkuch, 1999), the course has the aim to encourage participants to define a statistical problem and to pose statistical questions (First “P” in PPDAC), to plan and to prepare a data collection (Second “P” in PPDAC), to collect data (with regard to data management and cleaning – “D” in PPDAC), to analyze data (“A” in PPDAC) and to make conclusions of the data explorations (interpretation – “C” in PPDAC). We implemented cooperative learning environments like the “Think-pair-share” method to develop the statistical literacy and thinking components of our participants (see Roseth, Garfield & Ben-Zvi, 2008) and to support peer-learning, peer-feedback and expert-feedback. In this respect, in a “Think-pair-share” setting, students first deal with the task on their own (“think phase”), discuss about their findings in peers (“pair phase”) in a second step and finally discuss their findings in class with the teacher (“share phase”). All in all, the course consists of four modules: The first module deals with the generation of statistical questions, the preparation of data collection and the collection of data. Here the participants get to know the “PPD” elements of the PPDAC cycle with a special emphasis on the generation of adequate statistical questions (see Biehler, 2001). The second module has the intention to introduce the participants into data analysis with TinkerPlots. Here the participants can learn first steps in data analysis using data cards and hands-on-activities and then use TinkerPlots for first explorations in small datasets. Because the focus is on the “AC” (Analysis & Conclusion) elements of the PPDAC cycle, the datasets are given to the students in an already prepared form. Furthermore the students learn to describe and interpret distributions of categorical and numerical variables with special emphasis to the elements and characteristics of distributions like center, variation, etc. as it is proposed in Rossman, Chance & Lock (2001) and Biehler (2007a, 2007b). Module three builds on module two and covers advanced data analysis with TinkerPlots in large multivariate datasets. Here the learners are introduced into comparison of groups (Pfannkuch, 2007). A major aspect in module two and three is to enable the participants to explore datasets and make their own statistical investigations with their own statistical questions (for a typical task for a statistical investigation in this respect see Figure 1). Module four has the intention to introduce the participants into inferential statistics, especially into randomization tests with TinkerPlots (Frischemeier & Biehler, 2014). Further details on the course design and the lesson plans can be read in Frischemeier (2017).

Accompanying research of the course: Stepwise development of statistical literacy and thinking

The course was taught by the first and second author and consisted of 14 sessions, each session lasted 90 minutes. One major goal of the course was to develop the elementary preservice teachers’ statistical literacy and thinking components with TinkerPlots. Since statistical investigations of complex datasets are new to our participants, we want to evaluate the statistical investigations in the introductory stage (module 2) and see in which way the quality will improve over time in cooperative learning environments. Two major research questions arise: How is the quality of the statistical investigations in the intermediate steps in module two? How does the quality of the statistical investigations develop in process of module two?

Participants, task and data collection

All in all 22 elementary preservice teachers participated in the course. All of them attended the course “Elementary Statistics” as described in the introduction. As a typical task a multivariate dataset with an exercise sheet consisting of four subtasks was given. As example you can see the “KinderUni”-task in Figure 1, where the dataset “KinderUni” had to be explored. The KinderUni dataset, is a (non-random

sampled) dataset with 28 variables containing information about leisure time and school activities of 39 pupils in the area of Kassel, Germany. In the introductory phase of module two the idea was, that learners at first explore small multivariate datasets to get used to data explorations with TinkerPlots and then to explore larger datasets in module three. So when working on the “KinderUni” task, the participants are at first (subtask (i)) asked for a short description of the dataset to get familiar with it. The second part (ii) of the task was to generate an appropriate statistical question. This statistical question of subtask (ii) is the starting point for subtasks (iii) and (iv). In subtask (iii), the participants are asked to create suitable graphs with TinkerPlots, which allow answering the statistical question arisen in part (ii). In subtask (iv) the participants are supposed to describe and interpret the TinkerPlots graphs of (iii) and finally to answer the statistical question posed in (ii).

„KinderUni“ task

- (i) Explore the dataset “KinderUni”. Which variables are taken into account?
- (ii) Generate an appropriate statistical question!
- (iii) Create suitable graphs in TinkerPlots, which help you to answer the statistical question posed in (ii). Take also further explorations into consideration.
- (iv) Describe your TinkerPlots graphs created in (iii). Interpret them and summarize your findings in a statistical report.

Figure 1: Task “KinderUni” as typical statistical investigation task in module two

The participants worked in pairs of two on the “KinderUni” task. So all in all, we had 11 pairs, who remained constant all over the course. When working on the task “KinderUni”, the participants were asked to document the procedure of their statistical investigations in written form in Microsoft Word with the TinkerPlots graphs implemented. We collected all word documents from the “KinderUni” task. As mentioned above, one major idea of the course was to improve the quality of the statistical investigations by peer-feedback and expert-feedback. This happened with cooperative learning activities like “think-pair-share”. First, in the “think” phase, all pairs worked on the task on their own and produced the preliminary version of the task (preliminary version: V1). Then two pairs came together and discussed the products of their statistical investigations (not necessarily with the same questions) in peers in the “pair” phase with the goal to find improvements for the TinkerPlots Graphs, for the descriptions of the TinkerPlots Graphs, etc.. Finally after revising the documents after the “pair” phase (version after peer feedback: V2), as a last step, the revised documents were discussed in plenum with the first and second author. After this phase the participants were again asked to revise their products for a final version (version after peer- and expert-feedback: V3). So for our data analysis we have the documentations on the statistical investigations of the participants as preliminary version (V1, n=11 documents), as version after peer feedback (V2, n=11 documents) and finally as version after expert-feedback (V3, n=10 documents).

Methodology for data analysis and coding

Our main goal was to rate the quality of the statistical investigations by points. Due to the huge amount of data, we used qualitative content analysis (Mayring, 2010) for rating the quality of the subtasks. We decided to weigh the subtasks (ii), (iii) and (iv) with equally two points maximum since these tasks are fundamental for the statistical investigation. In subtask (i) only one point is given, since this is an introductory task and easier than subtasks (ii), (iii) and (iv).

For subtask (i) we expected a description of the dataset (number of cases and variables, description of variables). In our course we have set the norm to begin every statistical investigation with an introduction. So two codes are given: “subtask (i) done correctly” and “subtask (i) not done correctly”. If subtask (i) is done correctly, one point is given, if it is not done correctly no point is given for this subtask. Details given with examples can be found in Frischemeier (2017, p. 350).

To distinguish the quality of statistical questions posed in subtask (ii), we took into account the classification of Biehler (2001) in “one-variable-” and “two-variable-” questions from a deductive point of view. So we distinguished whether the questions take into account one variable (example: “What is the distribution of the variable height?” - variable: height) or two variables (example: “In which way do boys and girls differ in respect to the variable height?” - variables: gender and height). For questions taking into account only one variable one point is given as maximum, because the exploration coming out of questions containing one variable is easier than for questions taking into account two variables. For “two-variable-questions” a maximum of two points are given. In between we inductively identified different qualities of statistical questions: So there can be “one-variable-questions”, which have just “yes” or “no” as answer (example: “Do 60% of the pupils have a mobile phone?”) – rated with 0.5 points, whereas “one-variable-questions” in regard to a characteristic of a distribution (example: “How many pupils have a personal computer?”) – rated with 1 point - are a little bit more sophisticated. Also in the set of “two-variable-questions” we find different types: There are questions leading just to a “yes”/“no” – answer (example: “Is there a difference between boys and girls in their time spending on computer use?”), whereas other types of questions lead to working out differences between the distributions (example: “In which regard does the computer use differ between boys and girls?”). Questions of the first type are rated with one point, questions of the second type are rated with two points. There is also another type of “two-variable questions”, which we call “open and complex”-questions like “which differences exist between boys and girls in regard to their leisure time activities?” This type of “two-variable-question” is also rated with two points. In this course we have set the norm to try to pose statistical questions which aim at two variables. As an example for our rating in regard to subtask (ii) we take the question “How many kids have a way to school of 30 or more minutes?” of the pair Anne and Alice. We rated the question with one of two points, since it only covers one variable (“way to school”) and it is aimed at one characteristic of a distribution (“how many ...?”). Further details for the categorization of questions are given in Frischemeier (2017, p. 350).

For subtask (iii) an adequate TinkerPlots graph has to be created, which enables participants to answer the statistical question posed in (ii). Since we want our participants to focus on the distribution of the investigative variable and on the influence in regard to further variables, we have set the norm in our course that the icons should be stacked in TinkerPlots and further explorations (taking into account other variables) have to be made. If all three requirements (informative TinkerPlots graph, stacked dots and further explorations) are fulfilled, subtask (iii) is rated with the maximum of two points. Table 1 shows the several ratings for subtask (iii).

Informative TinkerPlots graph, stacked and further explorations	2 points
Informative TinkerPlots graph, stacked and no further explorations	1.5 points
Informative TinkerPlots graph, not stacked and further explorations	1.5 points
Informative TinkerPlots graph, not stacked and no further explorations	1 point
Non informative TinkerPlots graph/missing TinkerPlots graph	0 point

Table 1: Overview of ratings and their definitions of subtask (iii)

As an example for our rating with regard to subtask (iii) we take the pictogram (with stacked icons) of Anne and Alice in Figure 2. With this TinkerPlots graph they are able to answer their question (“How many kids have a way to school of 30 or more minutes?”) posed in subtask (ii). Since icons are stacked, but no

further explorations are made, this graph is rated with 1.5 of 2 points. Further details and examples on the ratings of subtask (iii) can be read in Frischemeier (2017, pp. 354).

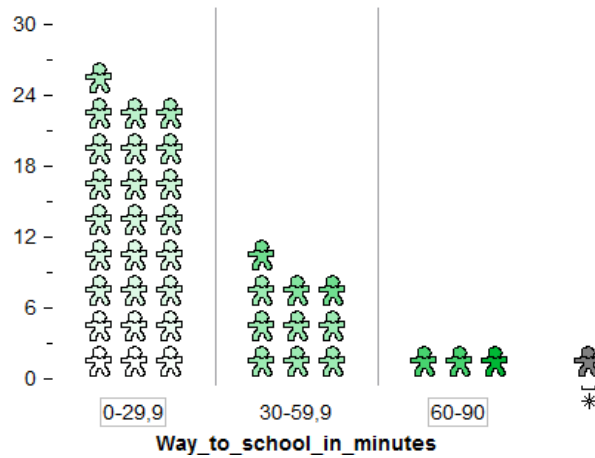


Figure 2: TinkerPlots graph for “KinderUni” task of Anne and Alice

In subtask (iv) the TinkerPlots graph (see Figure 2) has to be described adequately in at least one aspect and the question arisen in subtask (ii) has to be answered correctly. A maximum of two points are given, if both conditions are fulfilled. As adequate descriptions of the TinkerPlots graph we see elements like center, variation, shape, peaks, clusters and outliers (see Rossman et al. 2001, p. 48) but also absolute and relative frequencies of bins. For adequate elements to be carved out in group comparisons, see Frischemeier (2017, p. 42). In Table 2 we see the ratings for subtask (iv).

Component of TinkerPlots graph described and question (ii) answered correctly	2 points
Component of TinkerPlots graph described and question (ii) not answered correctly	1 point
Component of TinkerPlots graph not described and question (ii) answered correctly	1 point
Component of TinkerPlots graph not described and question (ii) not answered correctly	0 points

Table 2: Overview of ratings and their definitions of subtask (iv)

As an example for our rating in regard to subtask (iv) we have a look at the conclusion of Anne and Alice in subtask (iv): „We can see that 12+3 pupils have a way to school of 30 minutes or more.“ This was rated with the maximum of two points, since one component (absolute frequency of pupils in bins 30-59.9 and 60-90) of the graph is described and the question posed in (ii) is answered correctly. Further details on the ratings of subtask (iv) can be read in Frischemeier (2017, pp. 362).

Results

Let us have a look at the quality of the statistical investigations for the “KinderUni” task in module two in the different stages V1, V2, V3. For each team we rated the subtasks and calculated the success rate “points gained in all subtasks divided by the maximum points in all subtasks” for the “KinderUni” task in each stage (V1, V2, V3). In Figure 3 we see the distributions of the success rates in stages V1, V2 and V3.

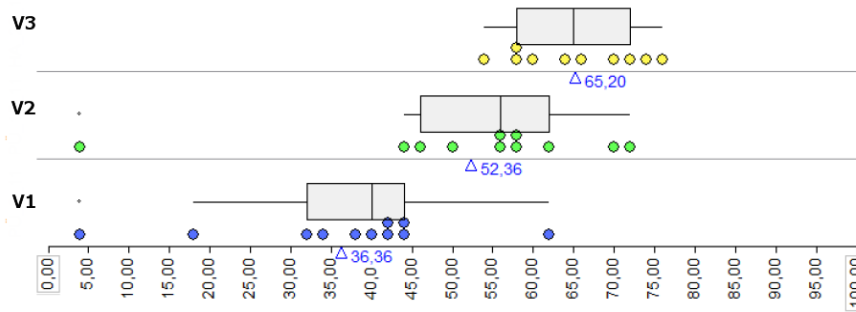


Figure 3: Success rates of the statistical investigations in stages V1, V2 and V3

Regarding to our research questions we can say that the median and also the mean (see blue triangles in Figure 3) of success rates of the different teams increase in the process of the several stages: In preliminary version (V1), where the pairs were on their own, 10 of 11 statistical investigations have a rate below 0.50, the median of the rates is 0.40, the mean of the rates is 0.3636. After the peer feedback phase (“pair”), there is a big positive shift in quality from V1 to V2. Exemplarily one peer feedback component which has often occurred was the advice to stack the dots in the plot to get a better view on the distribution of the data. The quality of the statistical investigations in V2 has increased a lot (mean=0.5236; median=0.56), since in this version only 3 of 11 statistical investigations are below the 0.50 rate. After the expert feedback (“share”) in stage V3 all reports are over the 0.50 rate, the median of the rates is 0.65 and the mean of the rates is 0.6520. The expert feedback concentrated most notably on prompts which suggest a better description of the TinkerPlots graph and a more adequate answer to the statistical question posed in subtask (ii). Finally we can identify a positive development of the quality from V1 to V3. We can also see that the distributions in Figure 3 are heterogeneous at the beginning (stage V1) and become more homogeneous in V2 and V3. For a more detailed look we will have a look at Table 3, which identifies the changes within the development of quality in between the four subtasks (i), (ii), (iii) and (iv) in the stages V1, V2, V3. We see that there is an improvement of quality in all subtasks, but the amount of the improvements differ on the kind of subtask. In subtask (i) there is a high quality (0.73) even in the beginning at the preliminary version (V1). This quality improves over time in the different stages V2 (0.91) and V3 (1.00). In V3, the rate is 1.00, which means that every pair began their report at this stage with an introduction of the dataset. In subtask (ii) we see the smallest development of quality: for the questions in the preliminary stage (V1) the rate is 0.47 on average, there is no improvement in V2 (0.47) and only a small improvement in V3 (0.50). Even at the stage V3, in subtask (ii) all questions were only rated with one point, since none of the questions went beyond single characteristics of a distribution or beyond “yes”/“no” answers. One reason might be that the peer and also the expert feedback concentrated too much on the improvement of subtasks (iii) and (iv) but not enough on the development of the quality of the statistical questions.

	V1	V2	V3
Average success rate in subtask (i) of all pairs	0.73	0.91	1.00
Average success rate in subtask (ii) of all teams	0.47	0.47	0.50
Average success rate in subtask (iii) of all teams	0.39	0.51	0.68
Average success rate in subtask (iv) of all teams	0.19	0.55	0.74

Table 3: Development of quality (average of rates) between the subtasks

In subtask (iii) the tasks were rated 0.39 on average at the preliminary stage and improved over time (0.51 at V2 and 0.68 at V3). In subtask (iv) the performance was very poor at the beginning (0.19), but improved in progress: In V2 the rate was 0.55 on average and in V3 the rate was 0.74 on average. So in summary we can say that peer feedback and expert feedback in a think-pair-share environment enhances a stepwise development of statistical literacy and thinking components with TinkerPlots. Especially with regard to subtask (iii) und (iv) the creation of TinkerPlots graphs and their description seem to improve after peer and expert feedback. Only in subtask (ii) problems with the generation of statistical questions occur and there was no “big” improvement of quality.

Discussion and implications

The quality of statistical investigations depends on the statistical question raised for the investigation. As we could see, some questions only lead to a short exploration because the answer to that question is just “yes” or “no”, whereas there can be also other questions which are aiming at carving out many differences between two or more variables. The analysis of the reports on the “KinderUni” task shows, that especially the creation of informative TinkerPlots graphs and also their description and interpretation with regard to the statistical question succeeds and the peer- and expert-feedback can improve the TinkerPlots graphs (subtask (iii)) and the descriptions and interpretations of the TinkerPlots graphs (subtask (iv)). The key point is the generation of adequate statistical questions aiming at more than only one variable. Although this was taught in our course, too many statistical questions lacked quality. For the re-design and the upcoming cycle of the course it would be important that there will be feedback on the statistical questions to improve their quality as well. Here it could be helpful to discuss adequate and non-adequate statistical questions in class to help learners to differentiate between adequate and non-adequate statistical questions.

References

- Arbeitskreis Stochastik der Gesellschaft für Didaktik der Mathematik. (2012). Empfehlungen für die Stochastikausbildung von Lehrkräften an Grundschulen. Retrieved 23.8.2016, from http://www.mathematik.uni-dortmund.de/ak-stoch/Empfehlungen_Stochastik_Grundschule.pdf
- Batanero, C., Burrill, G. & Reading, C. (2011). *Teaching statistics in school mathematics-challenges for teaching and teacher education*. Heidelberg/London/New York: Springer.
- Biehler, R. (2001). Statistische Kompetenz von Schülerinnen und Schülern - Konzepte und Ergebnisse empirischer Studien am Beispiel des Vergleichens empirischer Verteilungen. In M. Borovcnik, J. Engel, & D. Wickmann (Eds.), *Anregungen zum Stochastikunterricht* (pp. 97–114). Hildesheim: Franz Becker.
- Biehler, R. (2007a). Denken in Verteilungen - Vergleichen von Verteilungen. *Der Mathematikunterricht*, 53(3), 3–11.
- Biehler, R. (2007b). Students’ strategies of comparing distributions in an exploratory data analysis context. https://www.stat.auckland.ac.nz/~iase/publications/isi56/IPM37_Biehler.pdf
- Biehler, R., Ben-Zvi, D., Bakker, A., & Makar, K. (2013). Technology for enhancing statistical reasoning at the school level. In M. A. Clements, A. J. Bishop, C. Keitel-Kreidt, J. Kilpatrick, & F. K.-S. Leung (Eds.), *Third international handbook of mathematics education* (pp. 643–689). New York: Springer Science + Business Media.

- Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, 32(1), 9–13.
- Frischemeier, D. & Biehler, R. (2014). Design and exploratory evaluation of a learning trajectory leading to do randomization tests facilitated by TinkerPlots. In In B. Ubuz, Ç. Haser, & M. A. Mariotti (Eds.), *Proceedings of the Eighth Congress of the European Society for Research in Mathematics Education* (pp. 799–809). Ankara, Turkey: ERME.
- Frischemeier, D. (2017). *Statistisch denken und forschen lernen mit der Software TinkerPlots*. Wiesbaden: Springer Spektrum.
- 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: Cornelsen Verlag Scriptor.
- Konold, C., & Miller, C. (2011). *TinkerPlots 2.0*. Emeryville, CA: Key Curriculum Press.
- Mayring, P. (2010). *Qualitative Inhaltsanalyse: Grundlagen und Techniken*. Wiesbaden: Beltz.
- Pfannkuch, M. (2007). Year 11 Students' informal inferential reasoning: A case study about the interpretation of box plots. *International Electronic Journal of Mathematics Education*, 2(3), 149–167.
- Roseth, C. J., Garfield, J. B., & Ben-Zvi, D. (2008). Collaboration in learning and teaching statistics. *Journal of Statistics Education*, 16(1), 1–15.
- Rossmann, A. J., Chance, B. L., & Lock, R. H. (2001). *Workshop statistics: Discovery with data and Fathom*. Emeryville, CA: Key College Publishing.
- Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, 67(3), 223–265.