A LEARNING TRAJECTORY ON HYPOTHESIS TESTING WITH TINKERPLOTS – DESIGN AND EXPLORATORY EVALUATION

Susanne Podworny and Rolf Biehler
Institute of Mathematics, University of Paderborn, Germany
Podworny@math.upb.de

For a one semester course "Applied statistics – Understanding and solving complex problems via simulations with TinkerPlots" we designed and evaluated new learning material. The course follows an introductory course for statistics and probability and is meant as a focus on probability via simulation and inferential reasoning. At the end there is a four week learning trajectory on hypothesis testing and randomization tests with p-values. As new and innovative learning material for doing a simulation with TinkerPlots we developed a "simulation scheme", which should help students either to plan, structure or document a simulation and to force them to explain their simulation. Some of the learning trajectories and the simulation scheme will be discussed in detail.

INTRODUCTION

Simulations provide an opportunity to strengthen the understanding of statistical ideas (Konold, et al., 2007) and to support the learning process of learners when working on chance experiments (Maxara & Biehler, 2007). Our working group has developed a German localization of TinkerPlots 2 (not yet published) and is designing learning environments for data analysis and for probability with the use of TinkerPlots. There are some promising curricula using a simulation-based approach with Fathom™ for secondary level (Meyfarth, 2008; Prömmel, 2011) and with TinkerPlots for tertiary level (Zieffler, et al., 2013). For teaching probability we designed a course for pre-service teachers with the continuous use of TinkerPlots 2 to deepen the understanding of probabilities with a simulation-based approach to inferences. This course "Applied statistics – Understanding and solving complex problems via simulations with TinkerPlots" is leading from modeling basic chance experiments to inferential statistics including hypothesis testing and randomization tests. As innovative didactical material, the new developed simulation scheme is a major feature of the course.

INSTRUMENTAL APPROACH AND IDEAS FOR THE COURSE DESIGN

When looking at TinkerPlots as a software tool for learners, we can ask: How does the software TinkerPlots turn into a tool? Looking at the instrumental genesis process (Trouche, 2004), TinkerPlots turns into a tool when it is used by users for solving certain problems (Trouche, 2004, p. 282). There are two processes that represent the instrumental genesis: The *instrumentation* process is linked to the constraints and possibilities of a tool. This process is directed from the artifact to the subject and affects the subjects' use of the tool. The *instrumentalization* process reveals how the tool can be organized, structured and integrated into the subjects' practice. It is not always possible to clearly distinguish between those two processes, but both are involved in activities with the tool. A specific task may influence the way the tool is used by the user. "The instrument, then, is the psychological construct of the artifact together with the mental schemes the user develops for specific types of tasks. In such schemes, technical knowledge and domain-specific knowledge (in our case mathematical knowledge) are intertwined. Instrumental genesis, in short, involves the co-emergence of mental schemes and techniques for using the artifact, in which mathematical meanings and understandings are embedded" (Drijvers, et al., 2010, p. 1349). We developed a simulation scheme with the intention to support the *instrumentalization* and the *instrumentation* process and the mental schemes and techniques for using the tool. The scheme may be placed between the two processes.

The simulation scheme may help the subject to visualize, think of and express some of the constraints on a conscious level, which is why it is placed between the tool and the subject. Some of the instrumented action schemes can be displayed on the simulation scheme. The simulation scheme helps to connect and express the tool procedures within the mathematical world.

The simulation scheme may offer assistance at three levels (see Prömmel, 2011). The first is that before a simulation starts, the scheme may be used for planning the simulation. This

In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

involves thinking of the initial problem and considering the way of modeling the situation. The second level is assistance during a simulation as a structure for orientation as in the instrumentalization process. The simulation scheme shows which step comes next and can be used simultaneously when implementing a simulation in TinkerPlots. The third level involves documenting and reflecting on the simulation after it is conducted. This document may be used for in-class discussions about considering several ways of realizing a simulation or may be used for the next simulation as a worked example.

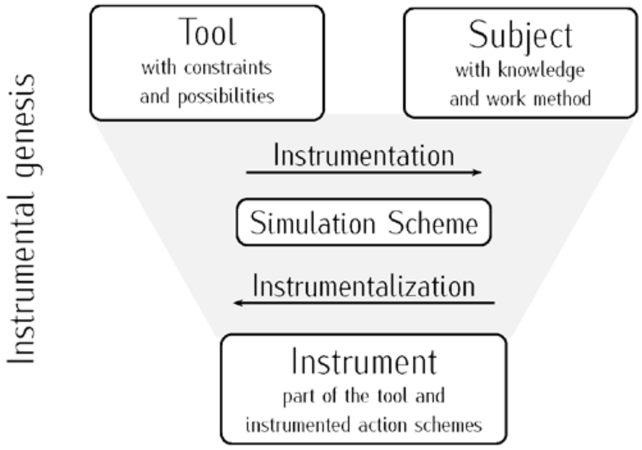

Figure 1: Adapted from Trouche, 2004, p. 289, Simulation scheme added

The design of the learning activities of the course followed the six principles of the "Statistical Reasoning Learning Environment" (SRLE) by Garfield & Ben-Zvi (2008). According to these principles the following design was chosen. Students work in pairs on most activities and talk and discuss ideas with their partner or with the pairs sitting next to them. Sequences of classroom discussion are embedded for supporting students' reasoning. Most tasks include real and motivating data or background to encourage the work. A main aspect of the design is the extensive weekly homework. This homework is used as an assessment "to monitor the development of their statistical learning as well as to evaluate instructional plans and progress" (Garfield & Ben-Zvi, 2008, p. 48). Regular reflections as part of a portfolio force students to deal with their understanding and their knowledge gaps. A few online surveys help to monitor the development of their statistical learning as well as to evaluate material. For further evaluation, a pre- and post-test was written to measure success of learning processes.

COURSE DESIGN AND CONTENT

At the University of Paderborn, pre-service teachers have to take a basic statistics course called "Elementary Statistics and Probability". In addition, pre-service teachers can take a seminar that goes deeper into the topics of this course in a subsequent term. From this point of view, our course was designed using modeling and simulations as an approach to inferential statistics. TinkerPlots allows easy access to simulate chance experiments via the *sampler*, which is why we chose this software. Furthermore, we think TinkerPlots is an adequate software for pre-service teachers at the primary or secondary level, and it fulfills very well the didactical requirements for software formulated by Biehler (1991, p. 190). These requirements are basically the definition of elementary models/machines with several types of elements; the possibility to combine these machines; options for further analysis of the sampled results and the 'run until' function.

The course includes 13 sessions (once a week for 90 minutes) with five topics. *1. Data analysis with TinkerPlots.* Our participants had already had basic experience with statistics. Therefore, the purpose of the first session is just to get to know the software and to refresh their knowledge on data analysis. Questions like "How can we analyze data with TinkerPlots?" and "What is a 'suitable' graph created by TinkerPlots for a given problem?" are discussed in this first session. *2. Basic simulations with TinkerPlots*. This theme lasts for four sessions and means learning on two levels. Students first learn how to simulate different situations with TinkerPlots and then learn how to connect their content and technological knowledge. During these sessions, we introduce the simulation scheme (see Figure 2) and use it from then on for supporting and guiding the simulation activities. Modeling different kinds of situations is the focus of this unit. The third theme is *3. Accuracy of simulations*, which lasts for two sessions and is more theoretical. We build on the law of large numbers and discuss the $1/\sqrt{n}$-law as a possibility to estimate the accuracy of a simulation and how it depends on the sample size n. *4. Independence and dependence* is the fourth theme, which lasts for two sessions. This is connected and realized with the metaphor of the sampler as a "data-factory" (Konold et al., 2007). A typical question is: "Does the simulation model fit the data?" The task is to change the model (which is constructed with the sampler) to see what happens to the data. The last theme, *5. Hypothesis testing with p-values and randomization tests,* includes four sessions and is a major theme in this course. We start here with a hands-on activity about identifying music either at mp3 or CD quality and introduce the idea of *p*-values in two sessions. During the last two sessions, we work on randomization tests (Rossman, 2008). The simulation scheme is used during most activities in all sessions.

SAMPLE ACTIVITIES FOR THE LEARNING TRAJECTORY ON *P*-VALUES

The learning trajectory on hypothesis testing is the final one in the course. *P*-values are commonly not used in German schools and are unknown to our students. The whole topic of hypothesis testing is known only to about half of our students. (They indicate this in one of the surveys). With regard to general misconceptions concerning *p*-values and hypothesis testing (see e. g. Garfield & Ben-Zvi, 2008; Vallecillos, 1994) and the general design principles described above, the topic is introduced slowly and with a hands-on activity (see sample activity 1) and lots of classroom discussion and explanations. After doing and discussing a homework task we finally introduce randomization tests as a possibility to visualize the random assignment process and to connect the design of a study with the inference procedure (see Rossman, 2008). TinkerPlots is a great support for the visualization process, especially for doing randomization tests.

Below, two activities are presented on the fifth theme: *Hypothesis testing with p-values and randomization tests*. The first activity introduces the subject and enables students to learn the logic of inference. The second activity is on a randomization test after a basic introduction to the topic.

Sample Activity 1: Audio Competition

This activity is based on an idea proposed by Riemer (2009) as a start for drawing inferences. It is about whether a person does better than random guessing in hearing the difference between a piece of music at mp3-quality (128 kbit/s) or at CD-quality (~1400 kbit/s). The starting point is the observation that a person got eight of twelve correct answers. There is a discussion about the fact that just guessing can produce results of eight, nine or even more correct answers. We ask students: "Without knowing something about the ability of a special person, do we think this person is a 'hearer' or not if they got eight (or nine or more) correct answers?" At this point students start a simulation for estimating the probability of getting at least eight (or more) correct answers when just guessing. This is followed by an introduction to hypothesis testing in a formal way with the steps *observation, null hypothesis, simulation, test statistics, p-value,* and *interpretation*. There are more aspects within the trajectory, which are discussed in pairs and in the whole group. For example, we ask: "How can we interpret a small *p*-value? Which conclusions are possible given a large *p*-value? What if the "guessing"-probability is not .5 but .8? What is then the probability of getting a prize?" All of these aspects are discussed and presented without calling the errors type I and type II but discussing some of their basic features. The "just guessing" or "randomly chosen" vocabulary is used from there on.

Sample Activity 3: Memorizing letters

This activity is based on an activity taken from an earlier version of Catalysts for Change (Zieffler, et al., 2013). In a previous session of the course students conducted their own study on 20 test persons memorizing letters (as homework, each student made about 20 people participate). They get two sequences of 30 letters each, arranged in groups of three or four letters. The groups of letters are in any order for sequence 1 and arranged in familiar abbreviations for sequence 2 (Example: Seq I: IDJ-FLO-MSQV-…; Seq II: JFK-IBM-…). They give one of these sequences to a test person and make them memorize the sequence. By repeating the letters, the test person gets a point for each correct letter until the first wrong letter (a score from zero to thirty points is possible). At the end we got data of n=592 test persons. Of course, the chosen people are not randomly sampled, and the assignment of the sequence to a test person is probably not random. We talk about this during a session. After introducing randomization tests to the students, they examine their own data (n=20) and compare the distributions of people memorizing sequence 1 and people memorizing sequence 2 by comparing the average scores. After analyzing the data, they examine the question of whether the average score depends on the sequence or not. At first they simulate the random assignment of a score independently of a sequence for their own study (n=20) and conduct a test to find a *p*-value. Afterwards, they take the data gathered by the whole class (n=592) and again carry out a randomization test. In the first small dataset (n=20) the *p*-values vary from .001 to .8, but with the large dataset of all the students (n=592), the *p*-value is near 0.00. There is plenty of room for discussion.

This activity is followed by three more tasks, where the supporting questions on the work sheets are gradually reduced until students only get a description of a situation and the question if something happens due to chance or for a reason. The simulation scheme is used for most activities within the learning trajectory mainly to make students think about their null model when starting the simulation.

THE SIMULATION SCHEME IN DETAIL

To support the whole process of a simulation with TinkerPlots we developed a simulation scheme with the goal to make students verbally describe all parts of their simulations. The scheme connects screenshots of TinkerPlots elements and a text box for each element.

Students may choose how to use the scheme: before the simulation as a "planning" tool, during the simulation as a support for the process, or after the simulation for documenting and reflecting the simulation. Figure 2 shows a completed scheme for activity 1 ("*Audio competition*") to the sub-question: How likely is it to get a prize (at least eight out of twelve correct answers) when someone is just guessing?

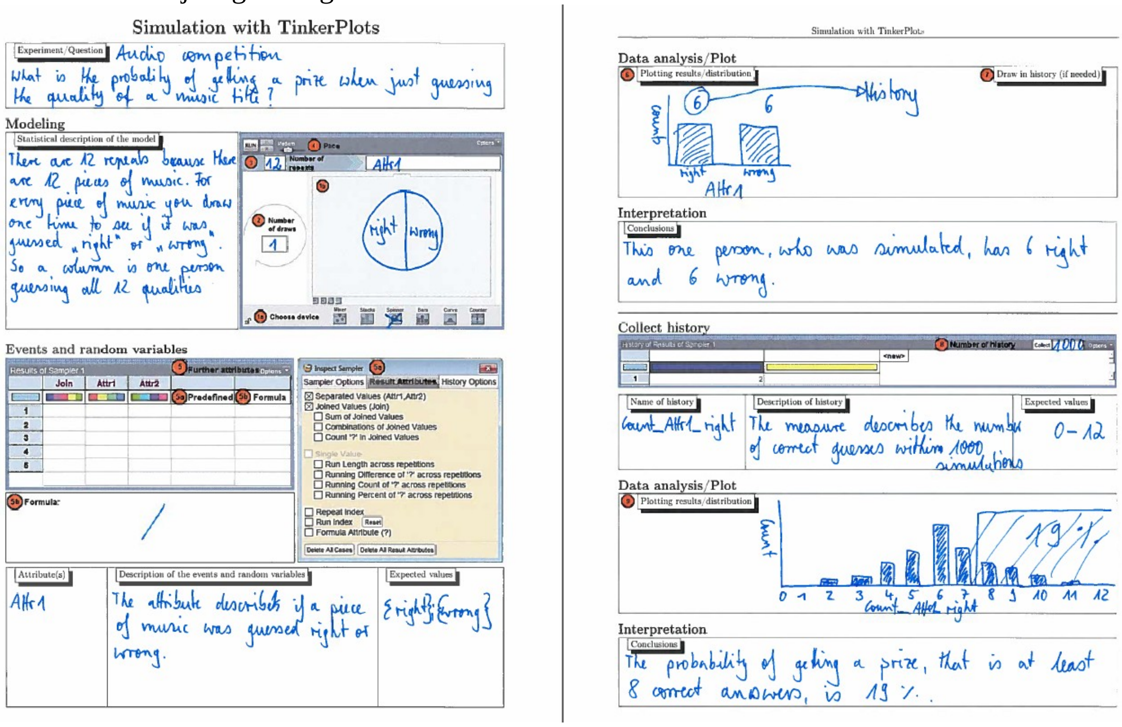

Figure 2: Filled-out simulation scheme for task "Audio competition"

The scheme is now explained in detail: What is the purpose and what should be learned in each step? Formally the scheme has eight main sections within several steps labeled with a running number. The last three sections are only needed if the simulation includes collecting history, otherwise the scheme ends at the middle of the second page. "Collecting history" is a function in

TinkerPlots to keep track of a measure when repeating a simulation; not every task needs to be modeled with this function. The sections are: Random experiment/question; Modeling; Events and Random variables; Analysis/Plot; Interpretation and additional for history: Collect history; Data analysis/Plot; Interpretation. The running numbers are a reminder of all the decisions the user has to make during a simulation.

The structure of the two-sided scheme is a guide to the simulation: It starts with a box for the question/random experiment that should be simulated. In figure 2 it is: *What is the probability of getting a small or a big prize when just guessing the quality of a music title?* In open tasks like in a hypothesis test situation, the question is sometimes not stated explicitly in one sentence. Therefore, formulating the random experiment and posing a single question for the simulation is the first step for students to dig into the experiment. Our experience shows that without a scheme, students start simulating without knowing what the question is that they want to answer with the simulation (see Maxara, 2009; Prömmel, 2011 for similar experiences). The need to formulate this in the simulation scheme helps them concentrate on the content. The next section of the scheme is a box near a screenshot of the sampler, labeled "Modeling". In Figure 2, a picture of the sampler is drafted with the explanation *There are 12 repeats because there are 12 pieces of music. For every piece of music, you draw it once to see if it was guessed "right" or "wrong". So a column [of the results table] is one person guessing all 12 qualities.* The sampler is the heart of the simulation, and most of the time there are several different options to model a random experiment with the sampler. The selected device for the simulation can be marked in the screenshot (no. 1a), how it was filled (no. 1b), how many draws were made (no. 2) and how many repeats (no. 3). There is space for renaming the attributes as well. In order to avoid forgetting to choose the speed of the simulation, this option gets the running number 4. The box next to the screenshot is labeled "Statistical description of the model". The explanation of the sampler as the model for the chance experiment must be written down here. Our students were always told to note if the model was with or without replacement because this can't be seen in the screenshot. The next section is for "Events and random variables" that are of interest for the task. The table and the window for Result attributes are plotted on the scheme (no. 5). Predefined attributes may be checked in the screenshot (no. 5a); other formulas may be written down in the "formula" box (no. 5b). It follows a small table where the attributes are named, and where a verbal description of them and their expected values can be recorded. In Figure 2 it is *Attr1 – The attribute describes if a piece of music was guessed right or wrong. – {right; wrong}.* The next page of the scheme starts with the section "Analysis/Plot" for simulated data. A rough outline as in Figure 2 of the graph with interesting summary statistics should be plotted (by hand) (no. 6). If histories should be collected, the corresponding value should be marked or encircled in this plot (no. 7).Whether the history was collected or not, the plot must be interpreted in the next box. In Figure 2 this interpretation is: *This one person who was simulated has 6 right and 6 wrong.* The last part of the scheme is for collecting history. The name(s) and the number of collected histories can be drawn in the screenshot (no. 8) and be explained in the following table, which is similar to the "Event and Random variable" table. In Figure 2 the number of "rights" was collected as history, labeled *Count_Attr1_right*, with the description: *The measure describes the number of correct guesses within 1000 simulations.* Expected values are *0-12*. Again the analysis of the data in the plot must be sketched in the following section (no. 9). The final interpretation of the results takes place in the last section. This is in Figure 2: *The probability of getting a small or a big prize, that is at least 8 correct answers, is 19 %.* Altogether the simulation scheme is a pen and pencil copy of the simulation enriched with explanations.

SOME SELECTED RESULTS

After the course we evaluated students' opinions concerning the simulation scheme (n=53 participants). They were forced to use the simulation scheme in most activities .Sometimes they had to use the scheme before doing something in TinkerPlots, sometimes they used it only during the simulation, and other times only after the simulation. Often they could choose at which point in their activity they use the scheme. Results of the survey show that most students don't want to use the simulation scheme before a simulation for planning purposes (73 % don't agree that the scheme is helpful for planning a simulation). Half of the students use it as an assistance during a simulation

(62 % agree that the scheme is helpful as a structure during the simulation) and many of them use the scheme for documenting a simulation (81 % agree that the simulation scheme is helpful for documenting a simulation). The simulation scheme is most commonly used as a worked example for a new task (82 % agree that a completed simulation scheme is helpful for a new task). Students were asked for an evaluation of the simulation scheme in an open response format. Here are some examples of their responses: "A good structure; I can reconstruct what I did in my simulation"; "The simulation scheme helps to reflect on my simulation afterwards", "I use the simulation scheme as an orientation for the process, but I take additional notes. I don't like to complete the scheme during the simulation, because this is confusing", "I think planning a simulation with the scheme is very tricky because I use the trial and error method when starting with the sampler. I don't really use the scheme at all". Students' attitudes concerning the simulation scheme are quite different, but encouraging for our further use of the scheme.

CONCLUSION

A simulation-based approach to inference requires special learning efforts from the students. They need not only statistical understanding, but modeling competencies, tool competencies and procedural knowledge of a hypothesis test. The simulation scheme as presented was well accepted and most students appreciated the support given by the scheme. But the simulation scheme should not be used for every task, like we did, because at some stage it bothers students to complete too many forms. It has to be decided when the scheme has fulfilled its scaffolding function. The additional learning effect seems to be small compared to the expense there is in completing the scheme for every task. One request was to include a box on the simulation scheme for expectations before the simulation starts. We are considering changing this in future research.

REFERENCES

Biehler, R. (1991). Computers in probability education. In: Kapadia, R. & Borovcnik, M. (Eds.). *Chance encounters: Probability in education* (pp. 169-212). Dordrecht: Kluwer.

Drijvers, P., Doorman, M., Boon, P., & van Gisbergen, S. (2009). Instrumental orchestration: Theory and practice. In V. Durand-Guerrier, S. Soury-Lavergne, & F. Arzarello (Eds.), *Proceedings of the sixth conference of the European Society for research in mathematics education (CERME-6).* Lyon, France.

Garfield, J., Ben-Zvi, D. (2008). *Developing students' statistical reasoning*. Springer Science + Business Media B. V.

Konold, C., Harradine, A., & Kazak, S. (2007). Understanding distributions by modeling them. *International Journal of Computers for Mathematical Learning 12,* 217-230.

Maxara, C., & Biehler, R. (2007). Constructing stochastic simulations with a computer tool– students' competencies and difficulties. In: *Proceedings of the fifth conference of the European Society for research in mathematics education (CERME-5).* Lamaca, Cyprus, 1-13.

Meyfarth, T. (2008). *Die Konzeption, Durchführung und Analyse eines simulationsintensiven Einstiegs in das Kurshalbjahr Stochastik der gymnasialen Oberstufe. Eine explorative Entwicklungsstudie*. Hildesheim: Franzbecker.

Prömmel, A (2013). *Das GESIM-Projekt. Rekonstruktion von Schülerwissen beim Einstieg in die Stochastik mit Simulationen*. Wiesbaden: Springer Spektrum.

Riemer, W. (2009). Soundcheck: CD contra MP3. Ein Hörtest als Einstieg in die Stochastik. In: *mathematik lehren 153*, 21-23.

Rossman, A. (2008). Reasoning about informal statistical inference: A statistician´s view. *Statistics Education Research Journal 7*(2), 5-19.

Vallecidos, A. (1994). *Theoretical and experimental study on errors and conceptions about hypothesis testing in university students*. Unpublished Ph.D., University of Granada, Spain.

Trouche, L. (2004). Managing the complexity of human/machine interactions in computerized learning environments: Guiding students' command process through instrumental orchestrations. *International Journal of Computers for Mathematical Learning, 9*(3), 281-307.

Zieffler, A., & Catalysts for Change. (2013). *Statistical Thinking. A simulation approach to uncertainty* (second edition). Minneapolis, MN: Catalyst Press.