HAL open science

Retrospective competence assessment in a PD course on teaching statistics with digital tools in upper secondary schools

Ralf Nieszporek, Rolf Biehler

To cite this version:

Ralf Nieszporek, Rolf Biehler. Retrospective competence assessment in a PD course on teaching statistics with digital tools in upper secondary schools. Eleventh Congress of the European Society for Research in Mathematics Education, Utrecht University, Feb 2019, Utrecht, Netherlands. hal-02422582

HAL Id: hal-02422582

https://hal.science/hal-02422582v1

Submitted on 22 Dec 2019

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.

Retrospective competence assessment in a PD course on teaching statistics with digital tools in upper secondary schools

Ralf Nieszporek and Rolf Biehler

Paderborn University, Germany;

Ralf.Nieszporek@math.upb.de; Biehler@math.upb.de

Recent changes in German national standards and curricula have made statistics including digital tools an inherent part of the final examination at upper secondary schools. In previous years, many teachers focused only on mandatory content like algebra or analysis and almost completely avoided optional stochastic content. Thus they are now facing the challenge to create and implement lessons for a topic in which they are inexperienced and feel insecure. The professional development course presented in this study has the aim to increase teachers' knowledge and competence in statistics, including the use of digital tools like graphic calculators and GeoGebra. A newly developed questionnaire was used to assess teachers' development during the course and led to preliminary findings concerning the knowledge and competence for advanced topics of inference statistics and the expertise of using digital tools in this field.

Keywords: Professional Development, retrospective competence self-assessment, probability and statistics, digital tools.

Design of the professional development course

There is a high demand for professional development courses on teaching statistics and the use of digital tools (like GeoGebra and graphic calculators) in Germany. Due to a recent change in the national standards (KMK, 2012), a new state curriculum was implemented in the German federal state of North Rhine-Westphalia (NRW) which made statistics including digital tools mandatory for the final exams (*Abitur*) in upper secondary schools. This obligation forced teachers to negotiate new and difficult challenges in their teaching practice. Many of them had not come into contact with statistics during their university education and they managed to also completely avoid this topic in their previous teaching practice. Therefore, content knowledge for teaching statistics or using digital tools in class is often missing (Batanero, Burrill & Reading, 2011). Teachers consequently feel insecure when teaching this topic. To meet the steadily increasing need for support, a professional development (PD) course was developed at Paderborn University. A team consisting of experienced facilitators, teacher educators, and practitioners revised an already existing course concept (Biehler, 2016) in the sense of *Design Research* (McKenney & Reeves, 2012). The improvements for the course were based on two foundations: on the one hand research on teacher education and proven ideas for designing teaching lessons (Biehler & Prömmel, 2010; Prömmel, 2013), and on the other hand concepts for PD course construction and research like the design principles of the German Center for Mathematics Teacher Education (DZLM) (Barzel & Biehler 2017) or the three-tetrahedron model for content-related PD research (Prediger, Leuders & Rösken-Winter, submitted, 2018).

The new state curriculum of NRW demands a wide variety of statistical topics to be covered in school lessons. It was possible to concentrate the essence of the curriculum in five PD attendance

days (see Table 1 for the content of each day) distributed over half a year with opportunities for teachers to try out the newly learned concepts in their lessons. In addition to that, the PD course also offers self-learning material so that teachers can deepen their knowledge on various topics like simulations with graphic calculators (GC), combinatorics, or the normal distribution.

Day	Topic	Content (examples)
1	Getting started in statistics and probability at upper secondary level, using GCs and simulation – in heterogeneous classrooms	Far-reaching and apposite examples for the law of large numbers and the $\frac{1}{\sqrt{n}}$-law, introduction to digital tools (GeoGebra, GC) and simulations
2	Conditional probability, statistics independence, and expectation as basic concepts in stochastic modeling	Natural (absolute) frequencies in tree diagrams, complementing contingency tables with double tree diagrams
3	Modeling with probability distributions, particularly with the binomial distribution: suggestions for teaching in a content-related and process-oriented way	The binomial distribution dynamically visualized with digital tools, interconnecting insights from previous day by using the same examples, strengthen modelling critique
4/5	Teaching inference statistics (hypothesis testing) with the goal of understanding, with authentic examples for the binomial distribution	Hypothesis testing via p-values to predefined significance level, errors of the first and second kind, power function, choice of the null hypothesis

Table 1: Topics covered in the five-day PD course

The aim of our study is to implement and analyze a PD course with an emphasis on increasing participants' knowledge and competence in teaching statistics with digital tools. This excerpt of our larger study focuses day five which covers the field of teaching more advanced topics of inference statistics testing like:

- Interpretation of hypothesis testing results (with focus on didactics and language)
- Type I and type II errors
- Choice of null hypothesis
- Operation characteristics and power function

More information on the PD design, PD aims and the course content can be found in Biehler, Griese & Nieszporek (to be submitted).

Theoretical framework

The TPACK framework created by Mishra and Koehler (2006) expands the mathematical knowledge for teaching framework (MKT) by Hill, Ball, and Schilling (2008). In addition to *content knowledge* (CK) and *pedagogical knowledge* (PK), Mishra and Koehler enriched the model with elements of *technology knowledge* (TK). However, those knowledge facets and their intersections like *pedagogical content knowledge* (PCK) or *technology pedagogical content knowledge* (TPACK) are not sufficient as the only input for a PD course. It is essential for teachers and for their daily practice to acquire additional skills beside pure knowledge on the topic, as they are confronted with different classes or situations and have to constantly undergo planning

activities. Planning activities can be performed with fictional classes where problems and difficulties have to be anticipated and adequate reactions must be found. Therefore, they can be implemented in PD courses rather easily to strengthen the skills of the participants. In contrast to knowledge dimensions these skills were grouped as competences in regard to Weinert (2001). Table 2 illustrates some of the knowledge and competence dimension addressed in our PD course.

Knowledge	Content	I have content knowledge relating to the topic.
	Background of the content	I know the didactical and curricular background and the learning goals of the topic
	Ideas for implementation	I know ideas for the implementation of the topic in lessons.
	Approaches for digital tools	I know approaches for using digital tools / GCs in the context of the topic.
Competences	Handling misconceptions	I am capable of recognizing and reacting appropriately towards misunderstandings (…).
	Planning	I can create a lesson plan incorporating the goals of the topic.
	Teaching the content	I can teach the topic in a goal-oriented way.
	Integrating digital tools	I am capable of implementing the approaches for the use of digital tools / GCs (…) in school lessons in a didactically advantageous way.

Table 2: Knowledge and competences addressed in the PD course, via statements to be rated

Teachers' heterogeneous knowledge and teaching experiences

For a long time, statistics and especially inference statistics have not been compulsory for upper secondary level, and were often neglected in university courses for teachers. This resulted in deficits regarding statistics knowledge of current in-service teachers, and some teachers might have mixed or mistaken content and horizon knowledge in some fields (Wassong & Biehler, 2010; Ball, Thames & Phelps, 2008). They are hardly able to simplify the content matter for class, which should also be taught in a PD course on statistics (Godino, Batanero & Font, 2007).

In addition to this deficiency, knowledge and competences towards digital tools are also demanded by new curriculum in NRW. GCs (or computer algebra systems, CAS) have also become mandatory in the final examination. Therefore, it is necessary to have at least a basic knowledge in programming, simulating or visualizing and interpreting distributions to fulfil the new demands. Due to the mismatch of teacher training and curriculum, this lack of experience must be addressed by a PD course, too.

Research questions

We will address the following research questions:

1. *General assessment of course success from participants' perspective.* To what extent (from before until after the PD course) do participants report to have acquired knowledge and competence in reference to conducting lessons in advanced topics of inference statistics testing?
2. *Suitability of PD course in respect to the heterogeneity of teachers' previous experiences.* What differences concerning the self-reported development of knowledge and competence

facets (from before until after the PD course) can be observed in different groups of teachers with comparable previous experience in teaching statistics and/or using graphic calculators?

Results for PD day 4 including a detailed description of the underlying theory, competence models and methodology can be found in Nieszporek, Griese & Biehler (to be submitted). The procedures used in this article form the foundation for this paper and the analysis of the course success. Due to space limitations it is only possible to give a brief overview of the results for PD day 5 in this article.

The ReCoS questionnaire

With reconsideration of the course concept and the corresponding theoretical background, the ReCoS questionnaire (see Figure 1) was developed to fulfill our needs. ReCoS is an abbreviation for **re**trospective **co**mpetence **s**elf-reports which describes the questionnaire quite well. It combines the knowledge and competence dimensions presented earlier (Table 2) with the specific content facets of each day. The PD course participants rate their before and after level of the facets subsequent to each PD day. The self-reports were made via German school grades (1=excellent,

Item No.:	I		II		III		IV		V		VI		VII		VIII	
	I have content knowledge relating to the topic XXX.		I know the didactical and curricular backgrounds and the learning goals of the topic XXX.		I know ideas for the implementation of the topic XXX in lessons.		I know approaches for using digital tools/GC in the context of the topic XXX.		I am capable of recognizing and reacting appropriately towards misunderstandings and pupils' faulty reasoning regarding topic XXX.		I can create a lesson plan which incorporates the goals of the topic XXX.		I can teach the topic XXX in a goal-oriented way.		I am capable of implementing the approaches for the use of digital tools/GC in context of the topic XXX in school lessons in a didactically advantageous way.	
	Before the PD	After the PD	Before the PD	After the PD	Before the PD	After the PD	Before the PD	After the PD	Before the PD	After the PD	Before the PD	After the PD	Before the PD	After the PD	Before the PD	After the PD
Interpretation of hypothesis testing results																
Type I and type II errors																

2=good, 3=satisfactory, 4=pass, 5=poor, 6=fail).

Figure 1: ReCoS questionnaire matrix, see Table 2 for the content of the columns

Pre/post or knowledge test designs were not feasible measurement tools for our project. Reasons range from organizational difficulties (lack of time for a detailed knowledge test due to a tight PD course schedule) to ethical concerns or motivation on behalf of the teachers. There are also good reasons for a retrospective test instrument like ReCoS. On the one hand different effects like *response shift* or *practice effects* were prevented. On the other hand, there was no *sensitization* towards a specific aspect or topic of the PD (Willson & Putman 1982).

Data collection and sample size

The ReCoS questionnaire matrix was adjusted to the topic of each PD day and distributed at the end of the respective day. A different questionnaire measuring individual experience levels via a six-point Likert scale (from *inexperienced* and *highly experienced*) was filled in by the participants at the beginning of the whole course. Individual anonymous codes were used to match the two questionnaires. 32 out of a total of 60 participants filled in both questionnaire types completely and thus only their data was used for further analysis.

Either SPSS 25 or R were used for statistical analysis like test for normal distribution (Kolmogorov-Smirnov), significance of competence increase (Wilcoxon signed-rank test) or building of scales of items and participants' groups.

Results

The general success of the PD course is illustrated in Figure 2. The ReCoS matrix shows the mean increase for each item. First of all, every change between the before and after competence self-assessment is positive. The changes were significant ($p<0.001$, Wilcoxon signed-rank test) and

Item No.:	I	II	III	IV	V	VI	VII	VIII
	I have content knowledge relating to the topic XXX.	I know the didactical and curricular backgrounds and the learning goals of the topic XXX.	I know ideas for the implementation of the topic XXX in lessons.	I know approaches for using digital tools/GC in the context of the topic XXX.	I am capable of recognizing and reacting appropriately towards misunderstandings and pupils' faulty reasoning regarding topic XXX.	I can create a lesson plan which incorporates the goals of the topic XXX.	I can teach the topic XXX in a goal-oriented way.	I am capable of implementing the approaches for the use of digital tools/GC in context of the topic XXX in school lessons in a didactically advantageous way.
	Differences between before and after score means							
Interpretation of hypothesis testing results	1.09	0.70	1.28	1.16	1.00	1.06	1.00	1.16
Type I and type II errors	1.09	0.80	1.25	1.32	0.97	1.09	1.06	1.22
Choice of null hypothesis	0.97	0.73	1.25	1.13	0.97	1.16	1.09	1.16
Operation characteristics and power function	1.61	1.18	1.64	1.52	1.22	1.48	1.30	1.41

were around 1 which is equal to an increase of one school grade.

Figure 2: Mean differences between the before and after scores of each item, a difference of one positive point describes an increase by one school grade, sample sizes between 29 and 32

Overall, the differences for each item do not vary much. Item II (*I know the didactical and curricular backgrounds and the learning goals of the topic*) is an exception. For almost every content facet the self-assessment increases less than 1 grade, which stands in contrast to the confidence in creating lesson plans (item VI) or teaching the topic (item VII). This might be a hint that teachers have insecure knowledge of the curriculum (a discussion of the curriculum was not part of the PD) or have difficulties understanding this specific item. Nevertheless, some scales within the questionnaire can be identified.

Content-specific facet	Knowledge	Competences	Technology
	α	α	α
Before scores			
Interpretation of hypothesis testing results	0.910	0.982	0.897
Type I and type II errors	0.924	0.948	0.909
Choice of null hypothesis	0.933	0.979	0.927
Operation characteristics and power function	0.942	0.969	0.910
After scores			
Interpretation of hypothesis testing results	0.913	0.965	0.848
Type I and type II errors	0.906	0.947	0.930
Choice of null hypothesis	0.901	0.968	0.951
Operation characteristics and power function	0.925	0.965	0.898

Table 3: Cronbach's α for the different scales, separately for the before and after items, sample sizes between 29 and 32

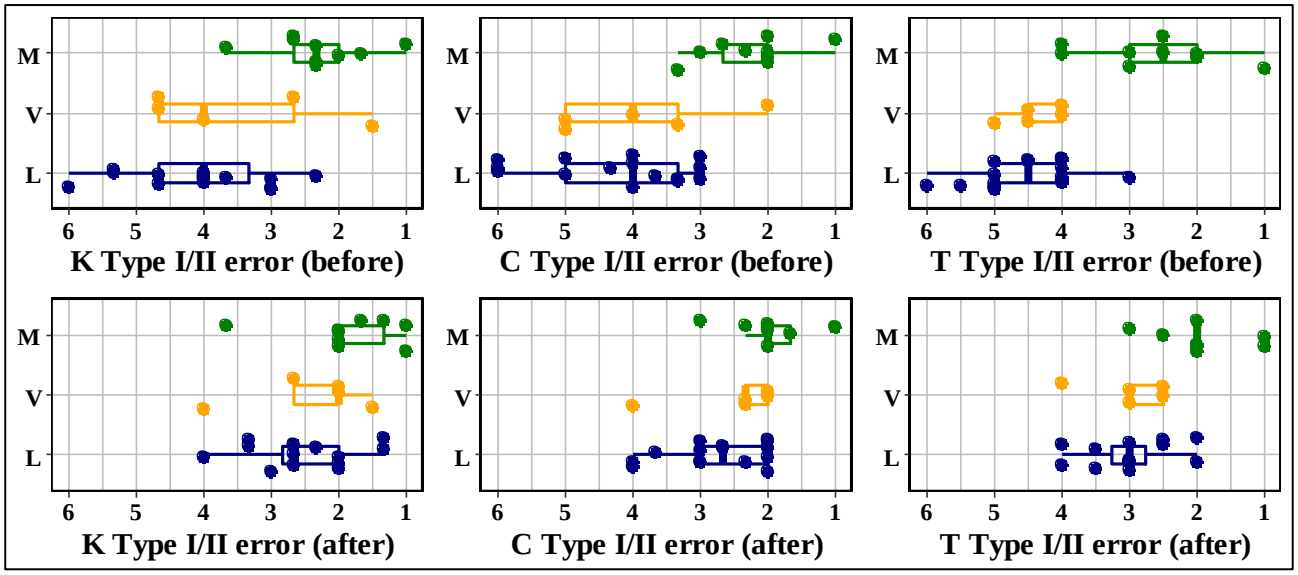
The theoretical framework of the course suggests a separation of the items into two scales *knowledge* (Item I – III) and *competences* (Item V – VII) for each content facet. Item IV and VIII form a separate scale called *Technology*, because technology plays a big role in the PD course and the abilities for using digital tools in school differ from the content components combined within *knowledge* and *competence*. The theoretical considerations are supported by empirical results. As is presented in Table 3, the identified scales within ReCoS show good to excellent Cronbach's α. In the following passages we will focus on the content facet *type I and type II errors.*

Table 4 shows the heterogeneity of experience among the teachers. The expertise reported in the field of teaching statistics and using the GC is not very high overall. It is striking that the participants were more unexperienced in handling digital tools (n=21), even when tools like GeoGebra or GC are commonly used for topics like algebra or analysis. This lack of experience for statistics appears to be a singularity.

	Experience in using GCs in statistics		
Experience in teaching statistics	Low	Medium	High
Low	n = 15 (L)	–	–
Medium	n = 5 (V)	n = 9 (M)	–
High	n = 1	n = 1	n = 1

Table 4: Distribution of participants in terms of self-reported experience in the use of GCs in statistics and in teaching inference statistics

For obvious reasons, we will disregard the groups consisting of only one person here, and instead focus on the participants who had low experience in teaching statistics and using GCs (*group L*), those who had varied experienced in both fields *(group V)* and those who reported medium



experience *(group M).* This selection reduced the sample size to 29 for further analysis.

Figure 3: Jitterplots and boxplots of the three scale pairs *knowledge* (K), *competences* (C) and *technology* (T) for the content *type I and type II errors* (Type I/II error), group L: blue (n=15), group V: orange (n=5), group M: green (n=9), scale range from 6=fail to 1=excellent, made with R

The increase of the mean differences (Figure 2) can be rediscovered in the boxplots in Figure 3. This figure visualizes the development of the three groups (L, V and M) for the content facet *type I and type II errors*. The before scores of each group fit their previous experience on teaching statistics. This means that the lowest experienced group (L) has the lowest before scores compared to the other two more experienced groups. Only group V, who also rated their experience in using digital tools low, shows equally low before scores in the technology scale. The after scores illustrate the huge increase of *knowledge*, *competence* and capability for *technology* reported earlier for the mean (Table 2). Even the most experienced group M rated themselves higher in all three scales after the PD course.

It is striking that there is a larger spread in the before scores of each group, which is remarkably reduced in the after scores of all three scales. Especially the participants of groups V and M rate their skills uniformly high. Only the lowest experienced group L shows a reduced but still high spread in the *knowledge* scale. There might be a ceiling effect at grade 2 (good) because only very few participants rated their skills higher with grade 1 (excellent). The rating system based on school grades might be the reason for that.

Discussion and remarks

Overall, the PD course increased the knowledge and competences of the participants by a significantly high amount (Figure 2) and was a success. But at the same time there are some hints as to where the PD course can be improved in a future design research cycle. We expected that the participants would rate their *knowledge* higher than their *competence* level after the PD course, because they supposedly would need to test the new material and ideas for teaching in their classes. There is a small difference between the score levels, but the *competence* after scores were almost as high as the corresponding *knowledge* scores. The different opportunities in the PD course to test and discuss the new input (for more details see Biehler et al., 2018) seem sufficient to encourage the participants to cover *type I and II errors* appropriately in class. Participants also feel more capable of using digital tools not only for calculating purposes but also for its didactical opportunities in the learning process.

The distribution of the before scores of each group (Figure 3) supports our division of the participants via their pre experience level. For example, the group of varied experience (V) shows medium high pre scores in *knowledge* and *competences* but has low scores in *technology*. The three different groups vary significantly from each other in the pre scores. This difference is still apparent in the after scores but relatively smaller since the after scores of groups L and V are almost on the same high level as group M. Also, the heterogeneity within each group is reduced on all scales.

It is important to have in mind that ReCoS only provides self-reported data and also that the sample size of 29 is not a large one. Therefore, we plan to contrast our results with data from other PD days and to use data from different other sources (questionnaires, audio transcript) to validate our findings. The comparison of the results from day 4 and 5 seems to be satisfying.

Acknowledgment

We would like to thank the Deutsche Telekom Stiftung who funds the DZLM and also the regional government of Arnsberg for cooperation and the opportunity to implement the course concept just as Birgit Griese for the help in the development process of the methodology and evaluation of day 4.

References

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education, 59*(5), 389–407.

Biehler, R. (2016). *Professional Development for teaching probability and inference statistics with digital tools at upper secondary level*. Paper presented at the International Congress on Mathematics Education, Hamburg. http://iase-web.org/documents/papers/icme13/ICME13_I1_Biehler.pdf last accessed 07/08/2018.

Barzel, B., & Biehler, R. (2017). Design principles and domains of knowledge for the professionalization of teachers and facilitators - Two examples from the DZLM for upper secondary teachers. In S. Zehetmeier, B. Rösken-Winter, D. Potari, & M. Ribeiro (Eds.), *Proceedings of the Third ERME Topic Conference on Mathematics Teaching, Resources and Teacher Professional Development (ETC3, October 5 to 7, 2016)* (pp. 16–34). Berlin: Humboldt-Universität zu Berlin. Retrieved from https://hal.archives-ouvertes.fr/ETC3, last accessed 03/12/2018.

Biehler, R., Griese, B., & Nieszporek, R. (submitted). Traditions and innovations for inference statistics at upper secondary level in Germany: An approach via a nationwide center for teacher training and professional development. *Journal for Mathematics Teacher Education.*

Biehler, R., & Prömmel, A. (2010). Developing students' computer-supported simulation and modelling competencies by means of carefully designed working environments. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July 11-16, 2010),* Ljubljana, Slovenia. Voorburg, The Netherlands: International Statistical Institute.

Godino, J. D., Batanero, C., & Font, V. (2007). The onto-semiotic approach to research in mathematics education. *ZDM, 39*(1), 127–135.

KMK. (2012). *Bildungsstandards im Fach Mathematik für die Allgemeine Hochschulreife (Beschluss der Kultusministerkonferenz vom 18.10.2012).* Retrieved from http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2012/2012_10_18-Bildungsstandards-Mathe-Abi.pdf, last accessed 07/08/2018.

McKenney, S. E., & Reeves, T. C. (2012). *Conducting Educational Design Research.* London, New York: Routledge.

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record, 108*(6), 1017–1054.

Nieszporek, R., Griese, B., & Biehler, R. (submitted). Development of competence patterns in PD courses for inference statistics at upper secondary level. *Journal for Mathematics Teacher Education*.

Nimon, K., Zigarmi, D., & Allen, J. (2011). Measures of program effectiveness based on retrospective pretest data: Are all created equal? *American Journal of Evaluation, 32*(1), 8–28.

Prediger, S., Leuders, T., & Rösken-Winter, B. (to be submitted, 2018). Which research can support PD facilitators? Strategies for content-related PD research in the Three-Tetrahedron Model. *Journal for Mathematics Teacher Education*.

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

Wassong, T., & Biehler, R. (2010). A model for teacher knowledge as a basis for online courses for professional development of statistics teachers. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics, ICOTS8*. Ljubljana, Slovenia; Voorburg, Netherlands: International Statistical Institute.

Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and selecting key competencies* (pp. 45–66). Göttingen: Hogrefe.

Willson, V. L., & Putnam, R. R. (1982). A meta–analysis of pretest sensitization effects in experimental design. *American Educational Research Journal*, *19*(2), 249–258.