DEVELOPMENTS OF TEACHERS' KNOWLEDGE FACETS IN TEACHING STATISTICS WITH DIGITAL TOOLS MEASURED WITH RETROSPECTIVE SELF-ASSESSMENT

Ralf Nieszporek, Rolf Biehler, Birgit Griese
Paderborn University, Germany
ralf.nieszporek@math.upb.de

Due to changes in the state curriculum of the German federal state of North Rhine-Westphalia, teaching statistics using graphic calculators, particularly for simulations, has become a focus of attention. This created a high demand for professional development (PD) courses in statistics. Considering this demand and approved concepts for learning and teaching statistics we created a PD course aimed at focusing on fundamental ideas and a pedagogically advantageous use of digital tools for simulations and interactive visualizations. Retrospective self-assessed questionnaires were used to measure the developments in participants' content, pedagogical, and technological knowledge. Our analysis shows interesting findings regarding heterogeneity in gaining new knowledge, for example in respect to handling pupils' misconceptions, didactically appropriate use of digital tools, or planning and conducting lessons.

CONTEXT OF THE PROFESSIONAL DEVELOPMENT COURSE

After the implementation of new national standards (KMK, 2012) in Germany and a new state curriculum in North Rhine-Westphalia (NRW), teaching probability and statistics became mandatory for the syllabus and for the final examination (Abitur). In addition, the use of digital tools like graphic calculators (GC) became mandatory in schools. Dealing with these changes was and still is a challenge for upper secondary school teachers and led to an enormous need for professional development (PD). In 2013, the Paderborn department of the German Center for Mathematics Teacher Education (Deutsches Zentrum für Lehrerbildung Mathematik, DZLM) consisting of experienced school teachers and researchers created a PD course called "stochastics compact" (Biehler, 2016) to address this need. In total there were six implementations of the PD course, starting in 2013. The presented research is related to the latest version of the PD course. This course took place in 2017 in cooperation with the regional school administration in Arnsberg and consisted of five one- day modules. Each module had a different topic, stretching from an introduction to upper secondary statistics and probability education with digital tools and simulations over conditional probability and stochastic independence to hypothesis testing with authentic examples. The focus in this article is on the module *introduction to hypothesis testing*.

THEORETICAL FRAMEWORK

Course Concept

The course concept is based on the standards and design principles of the DZLM (Barzel & Selter, 2015). The creators also considered different results from stochastics education research. This led to a concept that (among others) takes typical students' mistakes into consideration and considers teachers' difficulties when teaching probability and statistics. A frequent mistake is the misinterpretation of large and small p-values in the context of hypothesis testing. Therefore, the *introduction to hypothesis testing* starts with p-value testing followed by testing with predefined significance level. This sets the stage for understanding the meaning of the test results and of the p-values. The course also incorporates approaches, which had been successfully tested in university courses or experimental classrooms like designs for teaching hypothesis testing (Prömmel 2013) or Bayesian reasoning (Wassner, Biehler, Schweynoch & Martignon, 2004). The course developers also bore in mind how technology like GCs or other digital tools can support the teaching and learning of probability and statistics (Biehler, Ben-Zvi, Bakker, & Makar, 2013).

The modules were designed on the basis of "fundamental ideas" in statistics and probability (Burrill & Biehler, 2011; Biehler & Eichler, 2015). The module on the *introduction to hypothesis testing* is built up in four steps:

1. A comprehensible introduction to hypothesis testing with p-values
2. A comprehensible transition to hypothesis testing with predefined significance levels

In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018), Kyoto, Japan.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

3. Interpretation of test results with special emphasis on language sensitivity and typical mistakes
4. Hypothesis testing and of the mathematical modeling circle

Whereas these steps are related to classroom content, a second important design aspect is related to the facets of professional knowledge, which teachers should master when teaching this content. The underlying competence model is based on Ball's, Thames' and Phelps' (2008) work and its enhancement by Wassong (2017). The competence facets that were covered in the PD course and later partly measured in the questionnaire for the topic *an introduction to hypothesis testing* can be found in table 1.

Table 1. Competence facets which were covered in the PD course

Knowledge type	Content Knowledge (CK)	Knowledge of Content and Teaching (KCT)	Knowledge of Content and Students (KCS)	Special Content Knowledge (SCK)	Technological Content Knowledge (TCK)	Technological Pedagogical Content Knowledge (TPCK)
Example	Knowledge of fundamental ideas of hypothesis testing, its procedure, and validity	Knowledge of how to plan and implement a lesson for introducing hypothesis testing either with p-value or with predefined significance level	Knowledge of typical mistakes and language errors regarding the interpretation of test results	Knowledge of how to react adequately on pupils' reasoning on the meaning of small and large p-values, and of how to react adequately to statements on the validity of various interpretations	Knowledge of how to use digital tools for visualizations and calculations in the context of hypothesis testing	Knowledge of advantages and disadvantages of the use of digital tools and possible misconceptions regarding hypothesis testing

Questionnaire Concept

On the basis of this course concept a questionnaire was developed. The above-mentioned content facets are reflected in the **re**trospective **co**mpetence **s**elf-reports questionnaire (short form: ReCoS). Figure 1 shows the arrangement of the four content facets (rows) and eight competence facets (columns) creating the two-dimensional matrix of the ReCoS. The competences in the columns measure different dispositions referring to *situation-specific skills* and *performance*. This follows the model *competence viewed as a continuum* (Blömeke, Gustafsson & Shavelson, 2015). The teachers were asked to grade their knowledge before and after the course, from 1 (excellent) to 6 (inadequate) into the questionnaire.

	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-orientated 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
Hypothesis testing with p-value																
Hypothesis testing with predefined significance																

Figure 1. Excerpt of the *re*trospective *co*mpetence *s*elf-reports questionnaire (ReCoS) for the module *introduction to hypothesis testing.*

A knowledge test was not a feasible tool to measure the competence levels of the participants in the project. The reasons for this decision were various: besides organizational difficulties (lack of time for a detailed knowledge test due to a tight PD course schedule), acceptance on behalf of the teachers, ethical concerns were crucial and the lack of an adequate and valid test finalized this point. Therefore, self-assessment was used. To avoid inaccuracy effects in self-assessed data like response shift, carry over, or sensitization for a topic (see Wilson & Putman, 1982), a retrospective instrument was used. Other researchers successfully utilized retrospective measurements instead of a pre/post-test-design in other research contexts (e.g. Lam & Bengo, 2003, or Nimon, Zigarmi & Allen, 2010), which supports this decision.

RESEARCH QUESTIONS

The following research questions were evaluated. This paper focuses on the first two questions, but it is planned to present results for the other two at ICOTS10:

1. How did the self-reported competence level of the participants develop during the PD course?
2. What differences can be found in the developments of the self-reported competence levels of specific content facets of the module *introduction to hypothesis testing*?
3. What differences can be found in the developments of the competence levels of specific competence facets of the module *introduction to hypothesis testing*?
4. Are there any relations between participants' individual preconditions and level of motivation and their competence development?

DATA COLLECTION

The sample of the PD course consists of 60 upper secondary school teachers from two course locations. The questionnaire was distributed at the end of the module, therefore the participants knew about the content of each topic facet and the knowledge imparted when they were filling in the items.

RESULTS

There is no evidence for statistical differences between the two PD locations (mean differences: $p \geq 0.112$ Mann–Whitney U test). So the data was combined for further analysis. The self-assessed levels of each competence facet after the PD was significantly higher ($p \leq 0.001$, calculated with a Wilcoxon signed-rank test because normal distribution could not be assured) than before the PD. Every level of the content facet *hypothesis testing with p-values* before the PD was judged lower than those of others content facets. The increase (difference of grades after and before) of the competences relating to *hypothesis testing with p-values* was higher than those of the other facets (see Figure 2).

After the PD, the competence level was rated between grade 2 (good) and grade 3 (satisfying). There were no substantial differences between the competence facets of the different content facets, which is contrary to what was expected. It was anticipated that competences closer to classroom practices would be rated significantly lower, because of missing opportunities for field-testing. That was not the case and could be an indication for the effectiveness of the practice-related concept of the course.

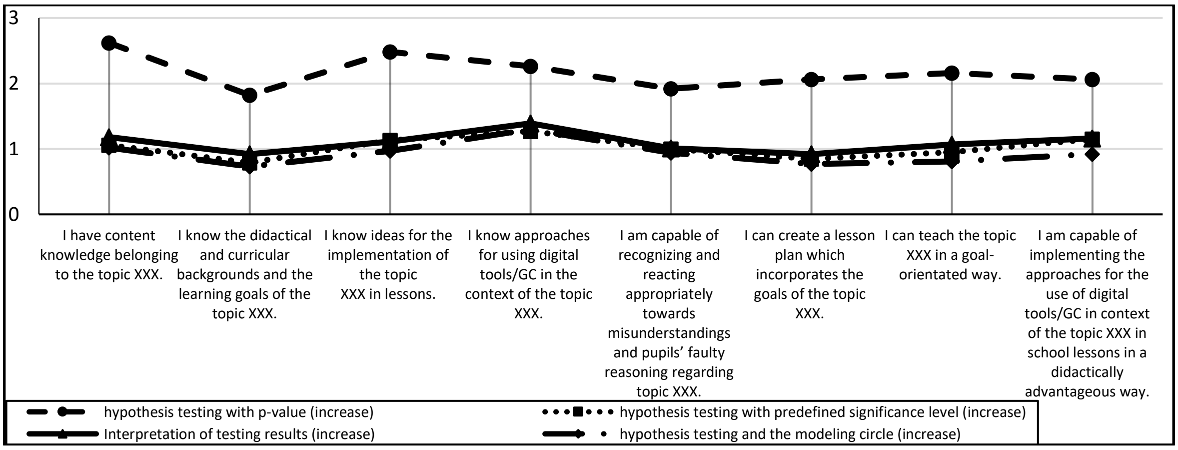

Figure 2. Increase of the self-assessments via ReCos from before compared to after the PD course, 1 point is equated to an increase of one grade

DISCUSSION AND REMARKS

Regarding the first research questions there is a recognizable subjective increase of the competence levels in all content and competence facets. As expected, the topic facets *hypothesis testing with p-values* was new to most of the participants, thus the teachers initially assessed themselves as rather low. Here, the PD course effected a substantial increase of the competence in these topic facets compared to all others. The increase compensated the initial differences in competence so that after the course the competences in all facets were almost identical.

For the presentation at ICOTS 10 it is planned to present a more detailed exploration of the questionnaire structure and the response behavior of the participants.

This article is an excerpt of a larger research project, which includes different kinds of questionnaires like assessment of learning targets (Frank & Kaduk, 2015) or *Stages of Concern* questionnaires (Biehler & Nieszporek, 2017). For future analysis, we intend to associate the results from these data with the results of ReCoS, aiming at new insights and potentially finding correlations between competence increase and personal requirements.

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., 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, J. Kilpatrick, & F. K. S. Leung (Eds.), *Third International Handbook of Mathematics Education* (pp. 643–689). New York: Springer.

Biehler, R. (2016). *Professional Development for teaching probabilty and inference statistics with digital tools at upper secondary level*. Paper presented at the 13th International Congress on Mathematics Education (ICME13), Hamburg.

Biehler, R., & Eichler, A. (2015). Leitidee Daten und Zufall. In W. Blum, S. Vogel, C. Drüke–Noe, & A. Roppelt (Eds.), *Bildungsstandards aktuell: Mathematik für die Sekundarstufe II* (pp. 70–80). Braunschweig: Diesterweg.

Biehler, R., & Nieszporek, R. (in preparation) Upper secondary teachers' stages of concern related to curricular innovations before and after a professional development course on teaching probability and statistics including the use of digital tools in *Proceedings of the Tenth Congress of the European Society for Research* in Dooley, T., Durand–Guerrier, V., & Gueudet, G. (Eds.), *Proceedings of the Tenth Congress of the European Society for Research in Mathematics Education* (CERME10, February 1 – 5, 2017). Dublin: DCU Institute of Education and ERME.

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 – A Joint ICMI/IASE Study: The 18th ICMI Study (pp. 57–69). Dordrecht: Springer.

Blömeke, S., Gustafsson, J. E. & Shavelson, R. J. (2015). Beyond dichotomies: Competence viewed as a continuum. *Zeitschrift für Psychologie / Journal of Psychology, 223(1),* 3–13.

Frank, A., & Kaduk, S. (2015). Lehrveranstaltungsevaluation als Ausgangspunkt für Reflexion und Veränderung. Teaching Analysis Poll (TAP) und Bielefelder Lernzielorientierte Evaluation (BiLOE). In: *Arbeitskreis Evaluation und Qualitätssicherung Berliner und Brandenburger*.

KMK. (2012). *Bildungsstandards im Fach Mathematik für die Allgemeine Hochschulreife (Beschluss der Kultusministerkonferenz vom 18.10.2012)*

Lam, T. C., & Bengo, P. (2003). A comparison of three retrospective self–reporting methods of measuring change in instructional practice. *American Journal of Evaluation, 24(1),* 65–80.

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.

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

Wassner, C., Biehler, R., Schweynoch, S., & Martignon, L. (2004). Authentisches Bewerten und Urteilen unter Unsicherheit – Arbeitsmaterialien und didaktische Kommentare für den Themenbereich "Bayessche Regel". In C. Wassner (Ed.), *Förderung Bayesianischen Denkens: kognitionspsychologische Grundlagen und didaktische Analysen*. Hildesheim: Franzbecker.

Wassong, T. (2017). *Datenanalyse in der Sekundarstufe I als Fortbildungsthema. Theoriegeleitete Konzeption und Evaluation einer Multiplikatorenqualifizierung*. Wiesbaden: Springer Fachmedien.

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.