metheval
friedrich schiller university of jena
department of psychology
chair of methodology and evaluation research
Universitätsprojekt **Lehrevaluation** (www.ule.uni-jena.de)
erik.sengewald@uni-jena.de · anja.vetterlein@uni-jena.de

"I like your lecture because I'm good." Can we trust the self-evaluation of competencies?

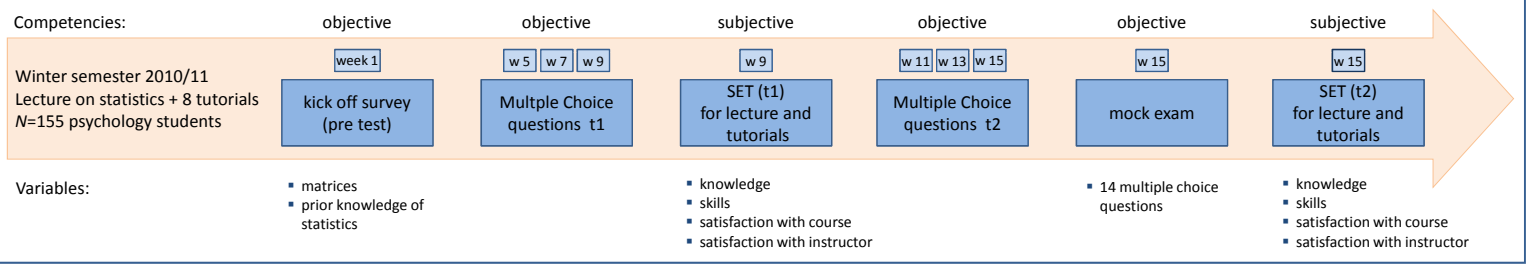
Erik Sengewald & Anja Vetterlein

Introduction

This study determines the relationship between objective measures of competencies and self-assessed knowledge in statistics. According to the title, this poster focuses especially how "students evaluation of teaching (SET)" and objective measures predict each other. Several studies discuss different approaches for this issue. Rodin & Rodin (1972) started to compare objective measurements in a mathematic course with the student rating of the course quality. They concluded, that "instructors are rated most highly in classes which have learned least" (Rodin & Rodin, 1972). Marsh (1975) criticizes the design of the Rodin study and uses the final exam as criterion for achievement. Accounting for the possibility that students compensate the lack of knowledge before attending the exam with a individual amount of learning, the present study uses a mock exam to explore the final achievement.

The self-assessment of achievement can be done with student-derived marks (see e.g. Rodin & Rodin, 1972 and Stefani, 1994). Another possibility is the overall satisfaction with the course as a quality-criterion (see Marsh & Roche, 1997). However, none of these SET measures cover a rating of acquired knowledge and skills. The current study combines subjective measures of knowledge and skills on latent dimensions with the objective performance of each student to examine the relationship between these measurements.

Design



Methods

The study accounts several times for subjective and objective measurements of competencies. We used a first semester course of statistics in psychology. In line with the schedule of the students, they had to attend one statistic lecture per week and one tutorial every second week.

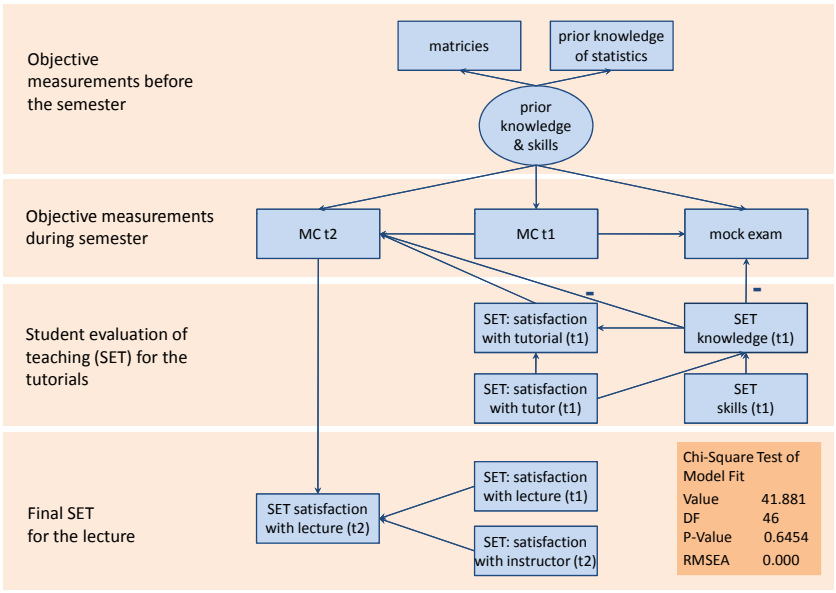
Prior knowledge: *N*=155 students were supposed to fill in a *kick-off survey* about their *prior knowledge in statistics*. Additionally 13 *matrices* were presented to get a general intelligence measure. These measurements were used as a pre test for skills and knowledge in statistics.

Objective measures: Each tutorial was held by one of four available female tutors and the students had to solve one *multiple choice (MC)* question in each tutorial. The final measure of knowledge took place in the last lecture of the semester by a written non-announced *mock exam*.

SET: In the middle of the semester the students were asked to fill in the questionnaire for *student evaluation of teaching* (PELVE inventory (e.g. Born, Loßnitzer & Schmidt, 2006)) and rate their subjective gained competencies and their satisfaction (t1) in the (a) lecture and in the (b) tutorial. In the last course session the evaluations of teaching were repeated (t2).

Data Analysis: We merged all data sources with R (R Development Core Team, 2008) by using one unique personal code of each student. The MC tasks were transformed into z-scores, so they are comparable over time. The students value on the latent SET-dimensions were assigned using Mplus (Muthén & Muthén, 1998-2007) and were included as manifest variables in the structural equation model.

Results



Discussion

The results replicates the finding, that teachers who receive better SETs are also the teachers from whom students learn the most (Marsh, 2007) on the student level within a course. Additionally, high achieving students rate the course better than others. However, further outstanding results are the negative partial regression coefficients of the self-assessed knowledge. Existing validation studies does not use such a measurement of acquired knowledge and compare it to objective performance measures (e.g. Marsh, 2007). Moreover, results of the current study indicate that students with a high pre test measure of knowledge tend to rate the achievement of knowledge via the SET lower than students with lower pre test values. These findings should not be interpreted as lack of validity at once, but used to question the utility of self-assessed knowledge as indicator of teaching effectiveness. Because "SETs [...] are difficult to validate, since no single criterion of effective teaching is sufficient" (Marsh, 2007) one should be careful using self-assessed knowledge as a criterion for teaching effectiveness.

Take home message

- SETs deliver valuable information about the learning process within a class.
- Satisfaction with the course increases with higher ratings on acquired skills, knowledge and higher satisfaction with the instructor.
- Satisfied students tend to achieve higher scores in objective measurements of knowledge.
- High achieving students are more satisfied with the course at the end of a semester.
- Structural equation modeling helps to show how SET ratings and objective measures predict each other (R^2: 27% - 56%).
- Analyzing change scores could answer the question whether the dimension of self-assessed knowledge is sensitive to measure change on an objective scale.
- Self-assessed competencies should be interpreted carefully when evaluating teaching effectiveness.
- The multisectional validation design of this study can be used to investigate the current questions on the course level.

References

Born, S., Loßnitzer, T. & Schmidt, B. (2006). Lehrveranstaltungsevaluation an der Friedrich-Schiller-Universität Jena – Eine Analyse der Dimensionalität der eingesetzten Fragebögen. In B. Krause & P. Methler (Hrsg.), *Empirische Evaluationsmethoden* (Bd. 10, S. 99 – 116). Berlin: ZeE Verlag.

Marsh, H. W. (2007). Students' evaluations of university teaching: Dimensionality, reliablity, validity, potential biases and usefulness. In R. P. Perry & J. C. Smart (Eds.), *The sholarship of teaching and learning in higher education: An evidencebased perspective* (pp. 319 – 383). Springer.

Marsh, H. W., Fleiner, H. & Thomas, C. S. (1975). Validity and usefulness of student evaluations of instructional quality. *Journal of Educational Psychology, 67*, 833 – 839.

Marsh, H. W. & Roche, L. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist, 53*, 1187 – 1197.

Muthén, L.K. and Muthén, B.O. (1998 – 2007). *Mplus User's Guide. Fifth Edition*. Los Angeles, CA: Muthén & Muthén

R Development Core Team (2008). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria.

Rodin, M. & Rodin, B. (1972). Student Evaluations of Teachers. *Science, 177*, 1164 – 1166.

Stefani, L. A. J. (1994). Peer, self and tutor assessment: Relative reliabilities. *Studies in Higher Education, 19,* 1, 69 – 75.