Teaching Quality in Higher Education

A Field Study Investigating Effects between Input, Process, and Output Variables Using Multiple Data Sources

Elisabeth Fischer

2019

Dissertation to obtain the degree Doktor der Philosophie (Dr. phil.)
at the Faculty of Human Sciences at the University of Kassel, Germany

Date of defence: 14th February 2019

Reviewers: Prof. Dr. Martin Hänze
Prof. Dr. Elke Heise
Prof. Dr. Birgit Spinath

doi: 10.17170/kobra-20191128820

Contents

Preface

When I decided – against my long-held and strong-felt conviction of certainly not pursuing a doctoral degree and writing another thesis – to go for my own dissertation project in 2012, it was for three reasons: Through having been involved in a research project in educational psychology, I had discovered that I actually enjoyed scientific work. In addition, I had grown to cherish the working environment and the people I collaborated with in the department of psychology at the University of Kassel. And thirdly, I had some pressing questions about teaching quality in higher education that I thought needed answering.

My interest in higher education teaching had been triggered and spurred on by two things: First, my own experiences as a university student. I had noticed large discrepancies between courses that were comparable in size, methodological approach, and topic, but still had utterly different effects on me as a student. While I had the impression of being deeply involved and learning a lot in one course, another left me disappointed and without any change in perspective or substantial learning outcomes. As I intuitively sensed these differences in quality of higher education teaching, I started developing my own hypotheses and theories on what might be critical factors for university teaching to be effective. But, of course, these were my very personal theories and I had no way of seeing if they were right…

Second, in 2011 I had joined a research project that aimed at improving mathematics teaching to freshman teacher students and evaluated the effectiveness of the measures taken in a two-cohort quasi-experimental design, called LIMA (Biehler et al., 2013). Despite the considerable effort put into the modification of the central lecture as well as into the improvement of the students' learning environment, no significant differences were found between the final exams of the two cohorts (Hänze, Fischer, Schreiber, Biehler, & Hochmuth, 2013). These rather devastating results made me wonder: If even these comprehensive, well-thought-out changes in teaching failed, what could impact the quality of teaching and lead to an enhancement of student learning at all?! What exactly was it that made teaching ultimately effective?

At the same time, I got the impression that not much was known about actual effectiveness of higher education teaching and that a lot of advice offered in institutionalised settings such as teacher training centres at universities lacked empirical foundations. In contrast to research on teaching and learning in primary and secondary school, which I felt to be a particularly lively and fruitful field of research in Germany, I did not notice comparable advances with respect to tertiary education.

A number of years after the idea was born, I now present a dissertation thesis carrying the bold title "Teaching Quality in Higher Education". Of course, countless books have been written on the quality of higher education teaching and there are academic journals solely dedicated to the examination of teaching in university (e.g., Active Learning in Higher Education, Teaching in Higher Education). So it is obvious that this thesis cannot in any way cover all the crucial aspects of university teaching, let alone elaborate on the construct of teaching quality in depth. (Again, there are books and journals on the mere meaning of the word.) However, it may not have been the wisest of all decisions, but scope and title were still chosen deliberately – again, for three, possibly not

wholly sensible reasons: 1) to trigger interest and, maybe, provoke a little (What??! Does she really think she found the holy grail of university teaching after generations of well-respected scholars in arduous legwork have merely been able to explain thin slices and single aspects of tertiary education? Who does she think she is??!); 2) to be able to cover a broad variety of issues in the complex web of teaching and learning in higher education (Yes, I guess, I belong to the group of doctoral candidates that would like to investigate half of everything and refuse to commit to one specific matter of inquiry only.); and 3) I was really predominantly interested in the "big questions". Searching for German studies on higher education teaching before starting my own project, I came across various studies that investigated very specific questions in very specific contexts, which raised doubts about their informative value and practical applicability in me. Hence, I decided to try to find effects and relations that might apply to a wide range of settings – in the most common course formats, across subject disciplines, under ordinary real-life conditions. Needless to say that this was another aim set too high…

Now, I will reveal that much in advance: You will not find the holy grail of teaching in my thesis. But it is my sincere hope that maybe you will find pieces of information and thoughts about single aspects of teaching quality in higher education on the following pages that you regard as interesting and worthy of consideration.

Summary

This dissertation addresses the issue of teaching quality in higher education: What exactly is quality of higher education teaching? How can it be determined? Which characteristics are decisive? To systematise the complexity of teaching and learning in higher education, a framework with various system levels and the dimensions of input, process, and output was introduced first. Aiming to research teaching and learning as they commonly occur in higher education, a field study with a longitudinal set-up was devised, which captured all three dimensions. On five measurement points in the course of one semester, data were provided by teachers, students, and expert observers. The sample comprised $N_{courses}$ = 80 lectures and seminars in various disciplines with their respective teachers and the enrolled students ($N_{students}$ = 5,765) at a mid-size public university in Germany. The study is constituted by three substudies on disparate aspects of higher education teaching, each pertaining to one dimension of teaching quality, respectively.

The first substudy focused on the input dimension and investigated the significance of teachers' personal value of teaching as well as of constructivist and transmissive teaching beliefs for actual teaching behaviour. The analyses revealed differential effects of the teacher characteristics on various aspects of teaching, such as the quality of instruction, student involvement, and rapport. Particularly the teacher's priority on teaching seemed to be beneficial.

The second substudy focused on the process dimension of higher education and inspected the effectiveness of distinct teaching methods. It compared teacher-guided and student-activating methods with regard to their influence on students' cognitive involvement and learning outcomes – subjective learning achievement, academic competencies, and interest. While teacher-guided methods were associated positively with student learning, student-activating methods tended to show negative effects.

The third substudy focused on the outcome dimension and scrutinised the measurement of teaching quality via student reports. The validity of student ratings of two teaching aspects – student involvement and rapport – was examined by comparing them to corresponding observer ratings. Four student and three teacher characteristics were tested as potential bias variables. Analyses indicated that the student characteristics did not impede the student ratings; teacher characteristics, however, partially showed undue relations to the student ratings of teaching.

All in all, the input, process, and output dimension of higher education teaching proved to be informative with regard to quality teaching. As a common theme, the importance of university teachers emerged in the synopsis of the central findings. The dissertation closes with reflections on quality teaching in higher education and remarks on possible implications for the practice field.

1 Introduction

The sound education of young professionals is vital for any society of the 21st century. In Germany, student numbers are on the rise: For the winter term in 2018 an all-time record of 2,867,500 students were enrolled at German higher education institutions (Statistisches Bundesamt, 2018a). Since 2013 public money in the range of almost 1% of the gross domestic product has been invested in higher education every year. For 2018 the respective figure was estimated at 31 billion Euros (Statistisches Bundesamt, 2018b). In view of the considerable amount of human and financial resources, the significance of an effective system of higher education becomes evident. The effectiveness of education, in turn, is closely connected to the matter of teaching. The degree to which university graduates comprehend the basic knowledge and scientific approach of their respective discipline, to which they are able to apply abstract subject matter to real-life situations and solve complex problems, to which they also learn to reason, to respect multiple perspectives, and to act responsibly largely depends on teaching (Shavelson, 2010). Thus, the relevance of higher education teaching quality is hardly deniable.

Empirical research on teaching in higher education dates back to the 1920s (McKeachie, 1990). As McKeachie illustrated, among the topics investigated early on were class size and teaching methods, with studies contrasting lectures and discussions, for example. Beyond mere instruction modes, the role of teachers as experts and authorities or as facilitators of learning became subject of inquiry as well. Soon, scholars' interest expanded to independent study and peer learning and to the evaluation of teaching through students. With the closing of the 20th century, further focal topics included the relation between teaching and technology and cognitive approaches to teaching in higher education. During these decades of exploration, scholars became aware of the importance of a number of environmental factors, such as the subject discipline, and realised that the effectiveness of many teaching aspects varied with respect to distinct learning outcomes or particular student groups. Overall, McKeachie (1990) painted quite an optimistic picture of past research on higher education teaching, attesting a high level of liveliness to the field with progress in theory, methods, and established knowledge, plus significant contributions for practical use.

By now, there seems to be a myriad of publications on higher education teaching, ranging from guidebooks with recommendations for various forms of university teaching (e.g., Fry, Ketteridge, & Marshall, 2009) to compilations of the current scientific discourse on different research questions (e.g., Perry & Smart, 2007). A literature search in PsychINFO (29.11.2018) with the terms "teaching" and "higher education or college or university or post secondary or postsecondary or tertiary" as obligatory components of the title rendered 1,866 hits, of which 1,142 were articles in peer-reviewed academic journals and 564 were dissertations.

In spite of the considerable advances in research that are mirrored by the great number of publications, Abrami, d'Apollonia, and Rosenfield (2007, p. 392) drew a rather critical résumé of the state of the field. They identified a number of open questions pertaining to the core of teaching and learning in higher education and pointed out the remaining uncertainty about the mechanisms of effectiveness in university teaching: For example, while it is quite clear that different teaching processes lead to varied products on the side of the learner, the specific nature of these causal

relationships is somewhat unclear. Apart from that, the impact of diverse student characteristics, course features, and circumstantial aspects on the processes of teaching and learning are still rather blurry. Similarly unsatisfied, Marsh (2007) denoted another deficit: “It is remarkable that after nearly a century of extensive research, there is apparently no general theory of college teaching” (p. 374).

While research on higher education teaching indeed looks back on a long history with very fruitful phases yielding a lot of useful information on teaching practice in many English speaking countries, the field in Germany is both “younger” and “thinner”: After a first rise in the 1970s, research on higher education teaching took a back seat for a long time. Innovation in higher education was mainly pursued by means of political and institutional reforms with a number of model projects and structural transformations of study programmes (Metz-Göckel, Kamphans, & Scholkmann, 2012). It was only in the beginning of the 21st century that teaching in higher education was put back on the agenda. This was largely instigated by the Bologna process (Bologna Declaration, 1999), which initiated the assimilation of European higher education in order to enable the mobility of students and staff and to increase comparability of educational degrees. Far-reaching measures were taken, such as a unified structure with bachelor’s and master’s degrees and a common system of credits (Floud, 2006). The new focus on learning outcomes, on competences as desirable educational products, and on measures of quality assurance via student evaluations and programme accreditation brought about a radical change in the German perspective on quality of tertiary education (Hopback, 2004; Rudinger, Krahn, & Rietz, 2008). Recommendations of the German Council of Science and Humanities (2007; 2008) indicate how the awareness of the importance of the actual teaching and learning processes in higher education grew. The upswing of higher education teaching in public attention was further reinforced by measures such as the contest of excellent teaching initiated by the Standing Conference of the Ministers of Education and Cultural Affairs and the Stifterverband in 2008.

Apart from these contextual factors, the scientific approach to higher education teaching in Germany has been a bit special, too: Whereas the international scholarship of teaching and learning in higher education has an empirical approach (see Perry & Smart, 2007), in Germany, the domain of so-called higher education didactics is rooted in pedagogy. The German pedagogy, however, traditionally holds a theoretical approach and has weaker relations to empirical methods. With new public funding (e.g., https://www.qualitaetspakt-lehre.de/) and the strong focus on the enhancement of teaching, more and more centres of higher education didactics were built up at universities offering counselling and training for university teachers. To date, a large body of guidebooks with hints and recommendations for successful teaching (e.g., Böss-Ostendorf & Senft, 2014; Johansen, Jung, Lexa, & Niekrenz, 2010; Winteler, 2011) and compilations of different approaches and programmes of academic teacher trainings at distinct universities (e.g., Auferkorte-Michaelis, Ladwig, & Stahr, 2010; Schmohr, Müller, & Philipp, 2018) has evolved from higher education didactics. While the ideas and reports often accrue from experiences and certainly contain inspiring thoughts, they often lack conclusive evidence as to how effective the single practices really are. In comparison to the advances regarding teacher training in higher education, empirical research of the field was clearly neglected (Metz-Göckel et al., 2012). Educational psychology in particular has hardly concerned itself with processes of teaching and learning in higher education (Spinath et al.,

2012). However, with the general upturn of higher education teaching as a matter of public policy and practical implementation during the past ten years, a broader empirical approach to higher education has developed as well. Here, the research funding of the Federal Ministry of Education and Research for the investigation of the professionalization of teaching staff or the assessment of student competences in higher education has certainly had an important impetus.[1]

This study is part of the reactions to the lack of psychological research on higher education in Germany. Building on the international body of research, it not only aimed at providing information about higher education teaching in Germany but also at advancing the field as a whole. With a primary interest in the processes of teaching and learning in class, the study adopted a narrow understanding of higher education teaching. While aspects like regular visiting hours, the availability of learning material and online tools, or extracurricular support like courses on scientific writing certainly present important elements of the educational set-up of a university, this study solely focused on the teaching occurring during course sessions. Within these confines, a fairly comprehensive approach was chosen, though: Rather than targeting one specific feature of university teaching, this thesis sought to throw some spotlights on selected, disparate aspects of teaching quality in higher education. That way, the complexity of the issue under investigation – the multitude of aspects involved at different stages of teaching and learning in higher education – could be emphasized. The broad approach further allowed for choosing different perspectives that possibly provide valuable insight into the matter and for working on diverse questions.

So, what is the outline of this endeavour? Firstly, the ensuing Chapter 2 will give a short introduction into the topic of quality of higher education teaching and present a framework model, which is used to systematise the complex network of teaching and learning in higher education. Thereafter, Chapter 3 provides an overview of the whole research project of this dissertation. Subsequently, Chapters 4 to 6 all address specific questions pertaining to higher education teaching: First, three teacher characteristics are examined with regard to their effect on various teaching aspects; second, the effects of two kinds of teaching methods on student learning are investigated; and third, the validity of students' teaching assessment is scrutinised. Chapter 7, finally, integrates the central findings of this study, reflects on the distinct concepts of quality, and concludes with brief thoughts on what to consider in future research and implications for the practice field.

[1] Hochschulforschung als Beitrag zur Professionalisierung der Hochschullehre, 22.11.2007 (https://www.bmbf.de/foerderungen/bekanntmachung-294.html), Entwicklung von Professionalität des pädagogischen Personals in Bildungseinrichtungen, 27.05.2008 (https://www.bmbf.de/foerderungen/bekanntmachung.php?B=347), Kompetenzmodellierung und Kompetenzerfassung im Hochschulsektor, 06.10.2010 (https://www.bmbf.de/foerderungen/bekanntmachung-587.html)

2 Quality of Higher Education Teaching

2.1 Theoretical Considerations

The term *quality* is not research specific, but frequently used in everyday language. As such it has various meanings. According to the Oxford English Dictionary (Quality, 2018), there is an important distinction to be made with respect to the meaning of the word: On the one hand, quality may simply refer to certain characteristics and the way someone or something is like; on the other hand, the word may imply a high standard or a degree of excellence, which can either be low or high. Whereas the first is neutral and purely descriptive, the second meaning is normative and bound to values and criterions that determine quality. Applying these semantic considerations to the topic of teaching, quality can thus refer to teaching structure, teaching approach, and the occurrence of specific aspects of teaching in a descriptive manner, or it can indicate that teaching is evaluated as being good or poor with respect to certain criteria. While both denotations offer valuable perspectives, the term is usually employed with its normative meaning in education research.

Pondering over the meaning of teaching quality in its normative sense, Fenstermacher and Richardson (2005) firstly distinguished between good and successful teaching. The concept of g*ood teaching* is linked to the task sense of teaching and implies that it meets high standards for subject matter content and methods of practice. The notion of *successful teaching*, in contrast, captures the achievement sense of the term and relates to the intended learning outcomes (p. 189). More precisely, the authors formulated:

> By good teaching we mean that the content taught accords with disciplinary standards of adequacy and completeness, and that the methods employed are age appropriate, morally defensible, and undertaken with the intention of enhancing the learner's competence with respect to the content studied (…). By successful teaching we mean that the learner actually acquires, to some reasonable and acceptable level of proficiency, what the teacher is engaged in teaching. (p. 191)

The two categories of teaching quality are not necessarily connected. There could be successful teaching with morally dodgy content (e.g., how to steal food without being caught) or even with proper subject matters but dubious methods (e.g., beating or drugging children for compliance). Likewise, good teaching does not necessarily lead to the intended learning outcomes. With that in mind, Fenstermacher and Richardson defined *quality teaching* as the combination of both good and successful teaching. However, for teaching to be good and successful at the same time, a number of conditions must be satisfied. A set of factors conducive to student learning, such as willingness and effort on the side of the learner, supportive surroundings and the mere opportunity to learn, have to be given in addition to appropriate teaching practices (Fenstermacher & Richardson, 2005).

Of course, the concepts of good, successful, and quality teaching lead to distinct approaches of appraisal, respectively: An assessment of the goodness of teaching may neglect the learning outcomes and focus on the activities of the teacher, examining how well they conform to the standards of practice. That is, the measurement pays heed to the way students are taught but not to whether they learn anything. Conversely, to discern successful teaching the learning outcomes need

to be assessed, whereas the teaching process is subordinate. If the goal is to additionally study the connection between the goodness of teaching and its success, it becomes furthermore necessary to consider the preconditions of the learners, the surroundings, and the availability and extent of learning opportunities, as they contribute to student learning aside from good teaching.

Next to measurement considerations, Fenstermacher and Richardson's (2005) model of quality teaching is only a framework, which still needs to be substantiated with specific criteria for goodness and success. The authors note, for example, that scholars and practitioners from different backgrounds with varying paradigms and priorities will naturally have different "takes" on what counts as good or successful teaching. Apart from different schools of thought within educational science, there are also the perspectives of the two parties directly involved in teaching. While university teachers may have their very own conception of quality teaching (e.g., Kember, 2000; Veiga-Simão, Flores, Barros, Fernandes, & Mesquita, 2015), which inevitably influences their teaching practice, students may have quite different ideas of what constitutes good teaching (e.g., Lee, Kim, & Chan, 2015; Moses, 1985; Parpala, Lindblom-Ylänne, & Rytkönen, 2011). For both teachers and students these views may furthermore vary between subject disciplines. Aside from the individual opinions of faculty and students within tertiary education, there is also disagreement on the general purpose of higher education and, hence, the indicators of quality in teaching on a superordinate level. In 2008, the German Council of Science and Humanities pointed out additional conceptions of quality in higher education held by different stakeholders: While university teachers might firstly seek to properly convey their subject along with its respective body of knowledge and scientific methodology, employers' priority was often an adequate preparation for professional work, and the public required graduates to contribute to the cultural, social, technological, and economic progress of the society (pp. 19-20). These diverse aims of higher education are also reflected in the respective laws: The German law on the framework for higher education activities (Hochschulrahmengesetz, 2017) reads in § 7: Higher education is to prepare the students for their professional work and to convey the necessary knowledge, competencies, and methods of the respective line of study, so that the students will be enabled to act responsibly in a free, democratic, and social constitutional state. To sum it up, there is a multitude of goals, opinions, and expectations pertaining to university teaching on various levels of generality, which lead to different criteria of quality teaching in higher education.

One general criterion of quality of higher education teaching is its effectivity and efficiency (German Council of Science and Humanities, 2008). Thus, providing information as to how teaching can be effective and most efficient may be considered one of the major contributions of higher education research. Corresponding to Fenstermacher and Richardson's (2005) approach to quality teaching, Abrami, d'Apollonia, and Rosenfield (2007) divided the research on effective teaching into three subsections with a product, a process, and a product-process definition of higher education teaching, respectively. Following the *product definition*, the desirable student outcomes include cognitive products (subject matter expertise, general cognitive skills like analytical thinking, meta-cognitive skills), affective products (attitudes, interest, academic self-concept, motivation to learn), and occasionally psychomotor skills or aesthetic appreciation. So, there is not a single product of effective teaching but many; the value attached to the distinct products varies, though. According to

the product definition, teaching is regarded as ineffective, if it does not lead to the desired outcomes. The *process definition* emphasizes the acts of teaching rather than the outcomes. Research conducted in this area may examine instructors' activities before and during teaching, that is, the preparation and the delivery of teaching including aspects like classroom activities, enthusiasm, or rapport with the students. Here again, it is not a single process to be evaluated but there are numerous aspects that may be considered as potentially relevant elements of teaching. Within this paradigm, ineffective teaching refers to a bad choice of teaching acts or poor enactment. Lastly, the *product-process definition* postulates that the instructors' activities before and during the teaching produce the positive changes in students. Thus, this line of research is interested in linking certain characteristics of the teaching process causally to certain cognitive or affective outcomes. Consequently, teaching is classified as effective, if a particular method, approach, material, or other can be shown to render student learning or another targeted result.

While the exploration of quality teaching by means of establishing links between teaching processes and the respective products is particularly promising with regard to both understanding the functioning of higher education and the proactive improvement of teaching, this approach is also quite challenging. Helmke, Rindermann, and Schrader (2008) emphasised the complexity of learning processes in higher education and the many factors affecting student learning. Besides the teaching, the students' individual preconditions, their social and family background as well as their personal study circumstances play a role. The effectivity of university teaching is further dependent on its institutional context – the subject discipline, the study programme with its examination regulations, and the material resources of the university. Apart from the vast network of influencing factors, research of teaching and learning in higher education is further complicated by the fact that learning processes are never linear or direct effects of the teaching provided; any teaching situation represents merely an opportunity to the students that they can engage with in diverse ways. While one student may listen to a lecturer's explanation and thereby understand a certain subject matter, another student may be less attentive and, thus, miss out on this chance to gain knowledge. In fact, the actual learning of the students can be quite independent from the way of teaching. Not only that a conscientious and highly motivated student might make the best of any educational setting; there are also circumstantial aspects like particularly strict or important examinations, which may be very effective in fostering learning. Furthermore, individual student characteristics and circumstantial variables may also interact with the instruction delivered by a teacher (Helmke et al., 2008, p. 153).

Having reviewed the different denotations of quality, the basic conception of quality teaching by Fenstermacher and Richardson (2005), the distinct research approaches to higher education teaching as well as the challenges that come with determining the effectiveness of teaching within the complex network of higher education – which understanding of teaching quality and which research approach are used in the present study?

To examine higher education teaching from diverse angles, this study adopted both the neutrally descriptive and the normative perspectives of quality. From the outset, the characteristics of instructors and the aspects of teaching were firstly inspected neutrally. Yet, the mere choice of variables was influenced by thoughts on their relevance for student learning. Thus, certain assumptions about teaching quality in its normative sense are mirrored in the selection of variables

already. When the normative perspective on quality was adopted, the definition of quality teaching by Fenstermacher and Richardson (2005) was regarded as a basis. However, the goodness of teaching in its basal sense – morally flawless subject matter and teaching methods – was largely taken for granted presuming that abusive teaching would hardly occur in public higher education. Therefore, the primary understanding of normative quality teaching within this study was strongly connected to its effectiveness, that is, its success. So, teaching is perceived as quality teaching in its normative sense, if it ultimately leads to desirable learning outcomes.

Investigating the effectiveness of teaching, the complexity of higher education has to be taken into account. To help untangle the many factors that influence teaching and learning in higher education, a model that systemises the environmental structure and the process components is presented in the following section.

2.2 Model of Higher Education Teaching Quality

The ecological model of Bronfenbrenner (1994) provides a valuable framework to capture the complexity of learning environments. As a reaction to restricted investigations of single phenomena within narrow confines, Bronfenbrenner pushed for the inclusion of surrounding variables into psychological research. He stressed that human development takes place through interactions with other persons, objects, and symbols in the immediate environment. These interactions in turn vary systematically in form, power, content, and direction depending on the characteristics of the person and on environmental factors. Bronfenbrenner proposed a differentiated conceptualisation of the environment with a set of nested structures, where each lower level system is a part of the system above. The lowest system level (microsystem) he defined as the activities and interpersonal relationships a person experiences directly – in the family, classroom, peer group, or the workplace. In this immediate surrounding, so-called proximal processes occur. These have been shown to exert the greatest impact on a person's development. The second system level (mesosystem) is formed by microsystems and accommodates, for example, the processes in the respective institutions. Exosystems on the third level only indirectly affect the processes individuals undergo, and macrosystems on the highest level, finally, embody the overarching patterns of micro-, meso-, and exosystems, such as their culture with common belief systems, bodies of knowledge, customs, life-styles, and opportunity structures. To capture time in its historical sense as a property of the surrounding environment, Bronfenbrenner further added a temporal dimension in form of a chronosystem to the model.

Transferring this general model to higher education, the microsystem may represent the teaching and learning processes within single courses; on meso-level the study programmes and subject disciplines with their specific structures and traditions (e.g., course formats, examination modes) could be located; the exosystem might be the higher education institution with its resources, and the macrosystem would represent the regional, political, and social context. An individual level can further be added below the microsystems to capture the characteristics and intrapersonal processes of single persons (Braun, Weiß, & Seidel, 2014). All the levels of higher education influence student learning; the lower the more directly the influence and the greater the impact.

Now, to depict teaching in higher education, it does not suffice to arrange variables and processes on their respective levels of influence; the process character with starting conditions, the actual teaching and learning and the corresponding outcomes needs to be captured as well. In educational science teaching has been conceived as a complex process with various stages and multiple factors affecting it early on. Applying that to higher education, Dunkin and Barnes (1986), for instance, spoke of context, presage, process, and product variables. Features of the environment, such as the size of the institution, discipline, curriculum, or broader, society and culture were termed context variables. Characteristics of teachers and students conceptualised as possible sources or explanations of teaching processes, which can thus be used to forecast classroom processes, were labelled presage variables. Processes comprised the whole of teaching behaviour, be it the presentation of content or the personal interaction with students. The effects and consequences, e.g., academic achievement, but also completion rates, students' attitudes, and evaluations of the course were summarised as product variables. The portrayal of teaching as a complex process was also taken up by educational scientists in Germany. Models of teaching in higher education that emphasize the process character of teaching and learning can be found, e.g., in Helmke, Rindermann, and Schrader (2008), Helmke and Schrader (2010) as well as Ulrich (2016). Here, the preconditions of teachers and students are referred to as input dimension, same as the resources, like rooms, technical devices, and learning material available in the university. The teaching delivered by an instructor and the learning taking place on the side of the students make up the process dimension. The knowledge gains, personality changes, and other targeted student outcomes are comprised in the output dimension.

While the three dimensions provide a helpful framework to illustrate the process character of teaching and learning and to localise variables that may impact or result from certain quality indicators within these processes, they also suit as distinct approaches to measure quality. Originally attempting to improve the evaluation of quality in medical care, Donabedian (1966/2005) suggested a three-pillar model of quality analysis and also drew on the dimensions of input, process, and output: First, the outcomes of processes serve as a criterion of quality (*output*). This approach has a lot of advantages; among other, the validity of the criteria is usually undisputed, plus they are often fairly concrete and measureable. But it also comes with limitations; most importantly, it may remain unclear how the products came about. Second, the goodness of the process itself may be an adequate point of measurement (*process*). This approach involves judgements on a number of aspects regarded as high standard or appropriate. Hence, it requires a lot of effort firstly determining these aspects and then specifying the values and standards to be used in assessment. Here, the estimates of quality are less stable and less final than those derived from the measurement of outcomes. While these two approaches remind of the previously mentioned definition of quality teaching (Fenstermacher & Richardson, 2005) and the respective research approaches (Abrami, d'Apollonia, & Rosenfield, 2007), Donabedian adds the input dimension as a third gateway to quality assessment: The setting or structure, as he labels it, which may include administrative requirements, the adequacy of facilities and equipment, the qualification of staff, but also personal preconditions that affect the process under investigation (*input*). The assumption justifying this approach is that quality starting conditions with favourable circumstances and high potential will enable and lead to

quality processes and outcomes. As an advantage, indicators in this category are often fairly concrete and well accessible; compared to outcome indicators it is quite a distal measurement, though. Reflecting on the significance of the different quality indicators, Donabedian remarked: “Outcomes, by and large, remain the ultimate validators of the effectiveness and quality” (1966/2005, p. 694).

Despite the fact that Donabedian (1966/2005) originally developed his model to assess quality of micro-level processes – the physician-patient interaction, to be exact – it is well suited for application to other levels of analysis as well. Therefore, inspired by Braun, Weiß, and Seidel (2014), both models were combined to form the framework for investigating quality teaching in this study (see Figure 2.1). First, it can help to systemize processes and their determinants within the realm of higher education. And second, it shows the distinct approaches to assess teaching quality.

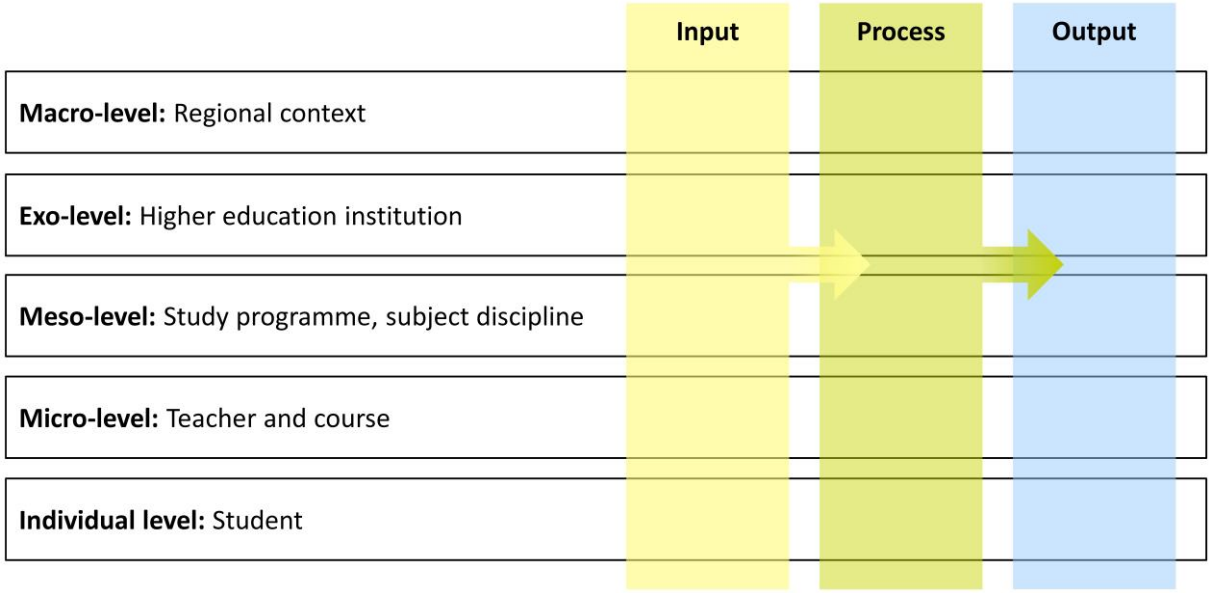

Figure 2.1. System levels of higher education with three process dimensions. Adapted from Figure 17.1 in Braun, Weiß, & Seidel, 2014, p. 435.

Braun, Weiß, and Seidel (2014) exemplified the system levels of higher education with respect to the situation in Germany: On macro-level the Bologna process (Bologna Declaration, 1999) has been particularly impactful during the first decade of the 21st century. Assimilating European higher education by installing two-step degree programmes with bachelor’s and master’s degrees and a common system of workload credits, the national administrations tried to enhance the mobility of students and staff as well as to assure the comparability of the degrees. In particular by setting the development of competencies as a primary goal of higher education, the policy process had a significant impact on teaching in Germany. Traditionally, input factors were trusted to ensure quality teaching in higher education. This was done by selecting highly qualified people and stressing a close, intrapersonal connection between research and teaching. The new focus on learning products resulted in a shift in attention towards outcome criteria. Now, teachers are explicitly asked to strive for the development of competencies on the side of the students. Apart from that, the

extensive establishment of measures of quality assurance such as the accreditation of study programmes was rather new to the German higher education system. Exemplary indicators for quality on the macro-level include on the input side the financial resources allocated to higher education; on the output side, low rates of college drop-outs and short study durations or the number of university degrees per age cohort can be used as indicators of quality (Braun et al., 2014).

Processes on the exo-level, that is, the single higher education institutions, are naturally influenced by macro-level dynamics, but also shaped individually by own characteristics – their reputation, tradition, or ties to local partners. With regard to teaching, the existence of institutions or staff units explicitly working on the enhancement of teaching quality in universities could be used as an indicator of quality within the process dimension (Braun et al., 2014).

The meso-level, which comprises subject disciplines and study programmes, influences teaching and learning processes more directly. The specific culture of a discipline manifests in rituals, traditions and structures, behavioural patterns, espoused beliefs, underlying assumptions, and embedded values, which are not conscious but guide actions (Umbach, 2007). Newcomers not only learn the content of the discipline, but also take in the language, the ways members interact with each other and the pedagogical techniques. Thus, it is not surprising that teaching varies systematically between disciplinary subcultures. Umbach (2007) reported differences with respect to the use of collaborative learning methods, the amount of higher order cognitive activities, or the requirement of student effort. The accreditation of study programmes, which assures a certain quality standard on an organisational level, poses a very powerful instrument of quality assurance on the exo-level. Quality indicators on meso-level include, e.g., the teacher-student ratios as an input factor, the way of teaching as determined by the respective traditions in the process dimension, and figures like the recruitment of young academics continuing the advancement of the subject (number of doctoral degrees) or the student drop-out rate on the output side. A popular measure of quality, which usually integrates indicators of all three dimensions of higher education teaching, are subject-specific and overall rankings of universities on national or even international level, e.g., the CHE ranking of German higher education institutions or the World University Rankings by the Times Higher Education.[2]

The primary focus when talking about teaching in higher education is, however, on the micro-level, where teaching actually occurs. And as soon as the students' learning is of interest, too, the individual level becomes equally important. As laid out above, the processes of teaching and learning in higher education cannot be analysed properly in isolation; factors like the subject discipline or the study curriculum directly affect teaching, as they determine requirements for taking the course, the learning content, or the examination regulations. Nonetheless, as it is the proximal processes in the immediate environment that exert the greatest influence (Bronfenbrenner, 1994), the variables and processes on these system levels are the most relevant ones to consider. Therefore, the subsequent section will zoom in on the lower levels of higher education and examine the variables and effects located here in more detail.

[2] Cf. https://ranking.zeit.de/che/de/ and https://www.timeshighereducation.com/world-university-rankings/2018/world-ranking#!/page/0/length/25/sort_by/rank/sort_order/asc/cols/stats

2.3 Input, Process, and Output Variables of Higher Education Teaching

The following overview of the variables pertaining to teaching in higher education does not cover the dimensions of input, process, and output comprehensively. Instead, some spotlights will be thrown on selected aspects that were considered relevant. Each construct is presented briefly with regard to their significance for teaching and learning in higher education. A compilation of the variables is given in Figure 2.2.

The chapter is firstly arranged in separate sections on the three dimensions of teaching. These are further divided according to the two lower system levels of higher education, the micro-level and the individual level. For reasons of simplification, the micro-level comprises both teachers and courses, even though one teacher may give various courses. The individual level refers to the students.

2.3.1 Input

Teachers and Courses

Thinking about input factors of higher education on the side of the teachers, the first aspect that may come to mind might be their *expertise* – concerning both the subject and the teaching task. Content knowledge and didactic or pedagogical competence both seem indispensable for quality teaching. As mentioned before, in Germany the primacy of a close connection between research and teaching is supposed to ensure a high expertise of university teachers. For that reason, faculty is usually required to simultaneously work in both areas (Braun et al., 2014, p. 444). But while subject expertise is built up throughout study and qualification times, there is no systematic or obligatory training of teaching staff at higher education institutions. In consequence, teaching usually is being learnt "on the fly", autodidactic, and rather accidentally (Helmke & Schrader, 2010). In comparison to many other European countries, Germany lacks behind as regards training programmes for higher education teachers (Winteler, 2006, p. 343). There is little research on the impact of varying levels of subject expertise. Usually, an adequate proficiency of university instructors is simply presumed. Also, it would prove very difficult to assess their highly specialised subject knowledge – at least if an actual test was to be used instead of such global measures as employment duration or professional status. Studies on the enhancement of teaching competences through training programmes, on the other hand, do exist. Dresel and Rindermann (2011) drew quite an optimistic picture of the effectiveness of training programmes reporting moderate to large effects of counselling on teaching quality. Stes, Coertjens, and Van Petegem (2013), in contrast, found no statistically significant effect of instructional development on teaching behaviour, neither for quantitative nor for qualitative student data. Synthesizing work (Penny & Coe, 2004; Stes, Min-Leliveld, Gijbels, & Van Petegem, 2010) suggested that, overall, teacher training, and thus enhanced pedagogical competence, seems to have positive effects on teaching, but that those effects are dependent on the specific counselling strategies used or the duration of the programme among other.

Apart from content knowledge and teaching competence, further teacher attributes that are deemed to be decisive are teachers' personal *teaching beliefs*, their subjective theories and conceptions of teaching. The topic of teaching beliefs triggered a lot of international research (e.g.,

Kane, Sandretto, & Heath, 2002; Kember, 1997; Trigwell, Prosser, & Waterhouse, 1999), and has meanwhile also been investigated in Germany (Lübeck, 2009). Instructors' varying conceptions of teaching are often described as a continuum with two poles: 1) teacher-centred teaching primarily understood as a means of knowledge transmission, and 2) student-centred teaching aiming at facilitating student learning (Kember, 1997; Kember & Gow, 1994; Samuelowicz & Bain, 1992). Studies trying to link teachers' personal beliefs to their teaching behaviour (e.g., Assen, Meijers, Otting, & Poell, 2016; Ebert-May et al., 2011; Martin, Prosser, Trigwell, Ramsden, & Benjamin, 2000; Mesa, Celis, & Lande, 2014; Prosser & Trigwell, 2006) to date remain inconclusive, though.

As a possibly relevant aspect of *motivation* the value teachers place on the teaching task has been inspected. Seeing that many university teachers tend to favour research-related activities (Cretchley et al., 2014), it has only been studied in connection to the wide-spread goal conflict between teaching and research, though. A meta-analysis by Hattie and Marsh (1996) indicated, however, that the quality of teaching and of research are essentially unrelated.

Of course, certain aspects of the teacher's personality, such as *enthusiasm* and *humour*, may also be significant. Triggered by the so-called Dr Fox effect (Naftulin, Ware, & Donnelly, 1973), teacher expressiveness and humour were studied extensively, revealing that expressiveness had substantial impact on global student course ratings and a small effect on achievement (Abrami, Leventhal, & Perry, 1982). More recently, Garner (2006) demonstrated congruently that students not only liked lectures with humorous insertions better, but also recalled more learning content afterwards.

The best researched course aspect is probably its size. The impact of *class size* was already studied very early on (McKeachie, 1990): The picture was not completely consistent, but results tended to show differences in student achievement in favour for small classes, which seemed to be dependent on the specific learning outcome measured. Yet, due to likely confounding with subject disciplines and the teaching methods employed, which affect student learning more directly, the variable's genuine significance must not be overrated.

Concerning input variables on the micro-level, there is a stable knowledge base in a few areas only, e.g., the nature of teachers' conceptions of teaching and a number of open questions with regard to others aspects, such as facets of teacher motivation. Numerous studies link teacher variables to student course evaluations or even measures of learning achievement (Abrami et al., 1982; Clayson & Sheffet, 2006; Shevlin, Banyard, Davies, & Griffiths, 2000) – often in the context of evaluation validity research; but little is known about the specific way certain teacher characteristics affect teaching.

Students

The individual characteristics of students and their respective significance for learning in higher education has been extensively researched. By now, several meta-analyses are available that summarize the numerous studies investigating student characteristics as determinants for success in higher education (e.g., Richardson, Abraham, & Bond, 2012; Robbins et al., 2004). Richardson, Abraham, and Bond (2012) collected findings from 7,167 articles published during the years of 1997 to 2010 with 241 samples for their meta-analysis on predictors of study success and assessed 50

distinct variables including demographic data, measures of cognitive abilities and prior achievement, personal traits variables, motivational constructs, learning strategies and study skills, and contextual factors such as social integration and financial support. Academic performance was indicated by the college grade point average.

Demographic variables and *study context* factors had virtually no correlations with achievement, a finding that is largely confirmed by Robbins et al. (2004). Among the *personality traits* Richardson et al. (2012) inspected conscientiousness and need of cognition showed small positive and procrastination small negative relations to achievement.

Indicators for the *cognitive abilities* of students are the high school grade point average (GPA), entry examinations, or intelligence measures. They showed medium correlations with learning achievement for the most part (Richardson et al., 2012). A meta-analysis of European studies reported that, in comparison, German final grades had the highest predictive validity for subsequent university grades. But, of course, the predictive power of cognitive abilities and prior knowledge as measured by school certificates does not solely depend on the country, but also on the subject discipline in tertiary education and the time span between school graduation and university studies (Trapmann, Hell, Weigand, & Schuler, 2007). Kuncel, Hezlett, and Ones (2001; 2004) confirmed the great significance of general cognitive ability, but pointed out the particular importance of subject-specific knowledge for study success, too. The predictive power of domain-specific parts of university entry tests proved to be bigger than the general verbal, quantitative, and analytic parts of the test, indicating that domain-specific prior knowledge is even more relevant in higher education.

Robbins et al. (2004) conducted a meta-analysis of 109 studies investigating tertiary students in the United States between 1984 and 2004, and focused on the role of students' *motivation* for persistence (retention) and performance (GPA) in college. Specifically, the investigated constructs comprised achievement motivation, academic goals, and academic self-efficacy, but also aspects like study skills, perceived social support, and social involvement. The persistence of students was related most strongly to study skills, self-efficacy, and academic goals, estimated true correlations with retention being even higher than for socioeconomic background, high school GPA, and scores of entry examinations. Performance, on the other hand, was best predicted by self-efficacy and second best by achievement motivation; here, high school GPA, and entry examination showed comparable predictive power. Congruently, self-efficacy and grade goal were found to be highly significant variables by Richardson et al. (2012) as well; goal orientations and other motivational variables only had small effects on performance, though.

The various measures of study skills inspected by Richardson, Abraham, and Bond (2012) mostly displayed small to medium-sized relations to the GPA, *effort regulation* appearing to be the most important one. Testing for variables' cumulative explanatory power over and above the cognitive ability measures of entry examination and high school GPA, effort regulation also turned out to be a particularly stable further predictor of college GPA. A longitudinal study over several years of study by Schiefele, Streblow, Ermgassen, and Moschner (2003) also found effort management to be the only aspect of learning behaviour that had a direct effect on academic achievement.

To capture students' dispositional attitude and approach towards learning, students' conception and style of learning (e.g., Busato, Prins, Elshout, & Hamaker, 2000; Van Rossum & Schenk, 1984) were studied. Busato et al. (2000) could not find positive associations between learning styles – meaning directed, reproduction directed, undirected, application directed – and academic success, though, the undirected style only showing a negative correlation with academic success. Closely related constructs are approaches to learning and learning strategies. As they are meant to inform about actual learning behaviour, they are discussed in the process dimension below.

In summary, we can say that distal variables like the social integration or financial support of students are of subordinate importance with regard to college performance, whereas cognitive factors are highly relevant for student learning. Motivational and volitional characteristics seem to be of medium significance and students' dispositional approaches to learning of small significance. While a lot of studies inform about the relations between student input factors and outcome indicators like the GPA, less is known about their impact on other desirable learning products such as the development of competencies or an increase in interest.

2.3.2 Process

Teachers and Courses

The process dimension on the course level of higher education refers to the actual teaching organised by the teacher, including aspects such as the delivery of subject matter and the interaction with students. It is the very heart of research on teaching (Dunkin & Barnes, 1986).

The bulk of information on quality teaching processes stems from analyses building on student assessment of teaching. Connecting the student ratings of certain teaching aspects in a concrete course with the respective measures of learning achievement is probably the most widespread approach to research higher education teaching and has led to meta-studies early on. The meta-analysis of Feldman (1989) was one of the most comprehensive works and is still cited as a good point of reference with regard to effective teaching (cf. Schneider & Preckel, 2017). According to Feldman's analyses, the *aspects of teaching* that have large effects on student learning are teacher preparation and organisation of the course, teacher clarity and comprehensibleness, and teacher stimulation of interest in the course and its subject matter. Furthermore, the encouragement of questions and discussion, the teacher's elocutionary skills, and his enthusiasm for subject and teaching, the clarity of course objectives and requirements, the quality and fairness of examinations, but also the teacher's availability and helpfulness, his sensitivity to and concern with class level and progress, and also intellectual challenge and encouragement of independent thought all have medium effects on student achievement. A teacher's feedback to students, his concern and respect for students and his friendliness also proved to be beneficial albeit to a smaller extent (Feldman, 1989).

In contrast to the high-inference approach usually inherent to student evaluation studies, Murray (2007) specifically focused on low-inference characteristics of teaching, *teaching behaviour*. Up to 100 distinct behaviours in the areas of clarity, expressiveness, interaction, organisation, speech quality etc. were assessed by external observers in three sessions per course. Correlations to student

ratings of overall teaching effectiveness were found, for example, for aspects of clarity – the use of concrete examples, stressing the most important points, and repeating difficult ideas. Similarly, behavioural aspects of organisation – putting an outline of the lecture on the blackboard, signalling the transition to the next topic, periodical summaries – and of rapport – offering to help students with problems, announcing to be available for consultation, showing concern for student progress – were associated to the student ratings of overall teaching effectiveness and overall course quality. Murray additionally investigated the behaviours' relations to other outcome measures as well: The three behavioural clusters of enthusiasm, task orientation, and use of class time were also significantly related to final exam grades. Rapport, clarity, enthusiasm, and task orientation further showed relations to student motivation. Although low-inference teaching behaviours were found to differ in frequency of occurrence depending on academic discipline or class size, their contribution to student-reported overall teaching effectiveness tended to be consistent across different contexts and situations (Murray, 2007).

One specific matter of higher education teaching has triggered particular interest and generated a lot of research and debate up to today: *teaching methods*. Springer, Stanne, and Donovan (1999) synapsed the existing research on small-group learning within the disciplines of science, mathematics, engineering, and technology. They found students who learned in small groups to demonstrate greater achievement than students who were instructed without cooperative or collaborative grouping. Moreover, students who worked in small groups were more persistent and expressed more favourable attitudes. Likewise, a lot of studies support the superiority of innovative, activating learning formats (e.g., Freeman et al., 2014). However, there are also findings in favour for traditional teaching methods (e.g., Struyven, Dochy, & Janssens, 2008) and studies reporting differential results with regard to the outcome measure (e.g., Dochy, Segers, Bossche, & Gijbels, 2003). Most meta-analyses on methodological aspects of teaching also report moderation variables, such as specific circumstances or certain modes of implementation that lead to the same instructional method having stronger or weaker effects (Schneider & Preckel, 2017). Due to this somewhat tangled picture, the debate on the effectiveness of teaching methods in higher education is alive and thriving, but precluding easy answers with regard to the use of teaching methods. In German higher education, traditional teaching still seems to be prevalent. Using video analyses, Seidel and Hoppert (2011) investigated seminars – courses with smaller student numbers, which allow for various teaching methods – and found them to be largely teacher-oriented, with no significant differences between subject disciplines.

Aside from teaching methods, the *technology* used in higher education teaching can be sorted to the process aspects. However, thus far tools like intelligent tutoring systems, animations, the simulation with virtual reality as well as blended and online learning had only small effects on student achievement at best (Schneider & Preckel, 2017).

A special part of teaching is the *assessment* of student learning. A number of particular approaches and techniques with their respective relations to student achievement are documented in Schneider and Preckel's (2017) overview. Linked to the topic of assessment the catch phrase of constructive alignment has been a prominent topic in the realm of higher education lately (Biggs, 2012). Constructive alignment refers to the harmonisation of learning goals with teaching methods

and the way of assessment. The intended learning outcomes are formulated as competencies on a certain level of proficiency. The teaching is designed in a way that provides opportunity for the learners to engage in activities needed to acquire and develop the targeted competencies; these are also assessed in the end. Research on this fundamental approach to teaching is still quite little, but, according to Biggs (2012), suggests positive effects for learning outcomes and student engagement.

Taking up a similarly superordinate perspective on learning environments in higher education, Vermunt (2003) identifies the appropriate *level of demand* as a key to quality student learning: To be effective teaching should firstly aim at high-quality learning, that is, the attainment of conceptual understanding, higher-order cognitive and meta-cognitive skills, e.g., with a strong focus on application. Secondly, teaching should adapt to students' increasing levels of competence, e.g., by increasing the complexity of the problems worked on, changing the required learning activities, or setting a higher bar for the learning objectives.

As can be seen from the great variety of variables mentioned, teaching processes in higher education are in themselves very complex. They are the central mystery of quality in higher education that needs to be solved in order to explain and improve university teaching. The process of teaching must, hence, remain at the core of inquiry in higher education research.

Students

The process dimension of higher education on individual level is concerned with the students' learning prompted by the teaching. Compared to aspects of higher education teaching, less is known about the processes of student learning. This may partly be due to the difficulties the examination of cognitive processes poses – unfortunately, no one can peek into a student's mind while he is taking in new information, reorganising concepts, or otherwise cognitively involved during a course.

One way educational scientists tried to tap student learning is via their *learning approaches* (Marton & Saljö, 1976) and self-regulatory *learning strategies* (Pintrich, 2004). Other than conceptions of learning or learning styles, which refer to relatively general and constant dispositions of learners, approaches to learning also depend on the specific learning situation and are thus more flexible (Gargallo López, Almerich Cerveró, Suárez Rodríguez, García Félix, & Garfella Esteban, 2013). Instruments of approaches to learning at least in part strive to capture the learning activities actually performed by students. Typically, scholars distinguish between two main approaches to studying in higher education: the deep approach aiming to understand the material and the surface approach aiming to memorise material. Partly, a third approach is included: the strategic approach aiming at obtaining the highest grades (Gargallo López et al., 2013; Richardson, 2005). While approaches chosen in a specific situation also depend on students' general conceptions of learning, research has shown that the same student may adopt different approaches depending on the concrete learning situation – the content, the demand, the quality of teaching and the nature of assessment (Richardson, 2005). Therefore, they are sometimes investigated as dependent variables of teachers' approaches to teaching (e.g., Kember & Gow, 1994; Trigwell et al., 1999). Self-regulatory learning strategies (Pintrich, 2004) are closely related to the construct of learning approaches, but capture both knowledge and application of certain learning activities. Rehearsal strategies include rote

learning through repetition, organisation comprises note taking or meaningful arrangement of learning content, elaborative strategies are, for instance, to summarise material in own words or to search for examples, and critical thinking may involve the questioning of the validity of information and materials. Furthermore, metacognitive strategies encompass techniques utilized to manage and oversee learning phases, including the tasks of planning and self-monitoring. Whereas critical thinking, elaboration, and meta-cognitive strategies were found to have small positive relations with college achievement, organisation and rehearsal learning were not associated with achievement. Equivalently, students' deep and strategic approaches to learning produced small positive relations with GPA, surface approaches showed slightly negative effects (Richardson et al., 2012).

Next to the students' learning-related approaches and strategies, their mere a*ttention* during university courses is an important process variable. Risko, Anderson, Sarwal, Engelhardt, and Kingstone (2012) have recently reproduced the finding that during a lecture mind wandering increases and memory for the teaching content diminishes. They could furthermore show a relation between lack of attention and the retention of the teaching content. Assessing student concentration during self-regulatory study, Richardson and colleagues (2012) report a small positive effect on the grade point average.

Surprisingly little research investigated active student *participation* in class. For example, no studies could be found on conducive effects of asking questions and other verbal contributions on student achievement. In contrast, the specific study technique of *note-taking* during teacher talks has been shown to have a weak positive effect on achievement. Moderation analysis revealed that it only unfolds its positive impact in the absence of presentation slides, though (Kobayashi, 2005).

As a last aspect of relevant student behaviour *class attendance* shall be mentioned. Credé, Roch, and Kieszczynka (2010) conducted a meta-analysis on the relationship between attendance and grades. They reported a strong association between students' attendance and their grade in a specific course, but also with regard to college grade point averages. Comparing student class attendance to individual input variables such as college entry tests, high school grade averages, or even study skills, it proved to be the best predictor for achievement. The attendance-grade relationship was slightly stronger for science courses than for non-science courses. In a ranking of 105 variables associated with achievement in higher education (Schneider & Preckel, 2017), frequent class attendance was placed on rank 6. Thus, this process variable of student learning can be regarded as particularly effective.

All in all, the processes of student learning have not been studied as comprehensively as, for example, student characteristics located in the input dimension here. Often, students' learning processes are inferred from their personal characteristics such as conscientiousness or effort regulation, but it may be valuable to also enhance knowledge on the actual learning activities of students. Research on learning approaches and strategies already shed some light on the different behaviours students' display in their study. However, comparably little research has investigated students' actual cognitive involvement in class.

2.3.3 Output

In this dimension, the two lower levels of higher education are particularly closely connected with each other. The outcome on course level is largely constituted by the outcomes of the single students. This applies in particular to student course evaluations. Student course evaluations are primarily meant to provide feedback to the teacher and are, hence, at the outset conceptualised on the micro-level. In contrast, students' learning outcomes are firstly relevant for each student individually. They are, therefore, covered in the second subchapter. However, seeing that any variable of student learning outcome if aggregated course-wise provides information on micro-level, too, collective measures of student learning products are mentioned here briefly as well.

Looking at different kinds of student outcome measures, Spinath and Brünken (2016, pp. 240-243) differentiate between objective and subjective measures of learning achievement. Objective measures include the knowledge gains and the development of competencies in a course. Subjective measures encompass the perceived learning achievement and the satisfaction with the way of teaching, the level of demand, and the specific circumstances.

Teachers and Courses

Student *course evaluations* provide teachers with information on the students' perception of their teaching and belong to the best researched aspects of higher education (e.g., Marsh, 1987; d'Apollonia & Abrami, 1997). They usually include the assessment of specific aspects like clarity of instruction, the organisation of a course, and the teacher's rapport with the students as well as ratings of global measures of overall teacher and course quality. The scientific discussion has mostly revolved around the issues of dimensionality (e.g., Abrami et al., 2007; Marsh, 2007) and in particular the validity (e.g., Marsh, 2007; Spooren, Brockx, & Mortelmans, 2013) of student assessment of teaching. The vast field of evaluation research shall not be discussed here in any detail, though (for an overview on the discussion on the validity of student evaluations, see Chapter 6). Even though findings with regard to the relation of course evaluations and measures of student achievement vary (e.g., Clayson, 2009; Cohen, 1981), student reports are widely perceived as a valuable indicator of quality teaching (Marsh, 2007).

Students' *collective learning outcome* can either be measured subjectively with self-report measures or objectively with tests or grades. The students' own rating of their learning achievement has been shown to correspond with objective achievement measures (Cohen, 1981), and may thus be considered a useful indicator. As Cohen conducted his analysis on student level, it is still unclear, whether this measure is meaningful between courses, though. The average examination grade or the passing rate of a course present objective output measures of teaching quality. Even so, comparing different courses these measures hardly provide any information on the actual amount of learning that has occurred in a course let alone the quality of teaching. The range and mean of the grades given in a course strongly depend on the respective teacher, the subject discipline, and even the institution. For example, final grades in German higher education were shown to scatter considerably between different subject disciplines; while 98% of the biology students received a good or very good grade, only 7% of the law students obtained grades better than satisfactory (German Council of Science and Humanities, 2012).

Thus, while learning outcomes may be regarded as the most important criterion for teaching quality, their measurement poses enormous challenges if courses covering different content are to be compared. Lacking fixed curriculums, which would allow for standardised tests in specific subjects, objective measures of learning outcome can only be used in laboratory or multisection studies as indicators for teaching quality. As soon as distinct courses are to be judged, subjective measures of teaching quality to date may present the best possible indicators of teaching quality on the micro-level of higher education.

Students

Output indicators for quality university teaching on individual level mostly pertain to the diverse products of student learning. Listing desirable effects of higher education teaching, Helmke and Schrader (2010) firstly mentioned the gaining of expertise, both fact and process knowledge, and the development of academic competencies such as self-regulated learning, discussion and rhetoric abilities, and communicative and cooperative abilities necessary for teamwork. Apart from that, further output criteria concerned student motivation and emotion, for example, the development of interest, of attitudes within the subject and towards research, and of moral standards. Similarly, Shavelson (2010) underlined the multitude of learning outcomes in higher education. He distinguished between domain knowledge, broad abilities, and soft skills. Domain knowledge comprised declarative, procedural and schematic knowledge as well as strategic knowledge on how to work and progress in the domain, while broad abilities included reasoning, decision making, problem solving, and communicating. According to Shavelson, soft skills encompass creativity, teamwork, and persistence as well as a sense of individual and social responsibility.

In a small, explorative study in Germany, Schrader and Helmke (2000) investigated the effectiveness of higher education teaching with regard to multiple criteria. They asked the students to what extend their study had supported them with respect to subject knowledge, practical abilities, social abilities, scientific thinking, working skills, general knowledge, autonomy, critical thinking, sense of responsibility, and personal development in general. Factor analysis revealed the two factors of cognitive and non-cognitive abilities. Trying to relate these two kinds of learning outcome to three measures of teaching quality – teacher engagement and rapport, the quality of the learning setting, and teaching quality – analyses showed consistently stronger relations for the cognitive learning outcome. So, possibly, cognitive learning products are influences by the teaching more directly.

The students' cognitive *learning achievement*, specifically their individual knowledge gain, can be indicated by both objective and subjective measures. Within single courses, examination grades present an objective measure, which informs students about their learning achievement. It contains information on both their personal mastering of the learning content in relation to the teacher's demand and – by means of social comparison – their position within the class. Of course, learning can also be assessed and reported by the students themselves. As mentioned above, the two measures of learning achievement were found to correspond (Cohen, 1981), supporting the validity of the students' appraisal.

The assessment of students' *academic competencies* as learning outcomes in higher education is a current topic of higher education research and has received special attention among scholars in Germany (e.g., Braun & Hannover, 2011; Braun & Leidner, 2009; Braun, Woodley, Richardson, & Leidner, 2012; Pant & Zlatkin-Troitschanskaia, 2016). Braun and Leidner (2009) reported moderate to strong relations between students' ratings of gains in competencies and their evaluation of teaching, but emphasised that the two constructs could still be distinguished empirically. Braun and Hannover (2011) could further show a connection between student reports of development of competencies and corresponding learning opportunities provided by the teachers in distinct courses. In contrast, *soft skills* (Shavelson, 2010) have hardly been investigated as an outcome of higher education yet. Shavelson suspected, this neglect in both research and public education might largely stem from the fact that it may be the most difficult to assess (p. 18).

Next to cognitive learning gains, student *motivation* is influenced by university teaching, too. For example, Patrick, Hisley, and Kempler (2000) could show that students' intrinsic motivation with regard to the learning material was boosted by teacher enthusiasm. Another relevant consequence of teaching in higher education is the students' *satisfaction* with a course. In a study with psychology students in Germany, Heise, Hasselhorn, and Hager (2003) investigated predictors for course satisfaction. Their analyses suggested that students' personal satisfaction with the course might be more closely dependent on aspects of teaching quality than objective indicators of student learning. Blüthmann (2012) confirmed the relevance of quality teaching for student satisfaction, but showed that additionally student characteristics such as their learning motivation and strategies contributed to the sense of satisfaction. As important affective indicators, students' motivation and satisfaction are at times assessed in addition to achievement to evaluate different learning environments (e.g., Hancock, Bray, & Nason, 2002; Krumboltz & Farquhar, 1957; Lim, Kim, Chen, & Ryder, 2008; Struyven et al., 2008).

At the end of the day, it is the outcomes of the single students, the whole higher education system ultimately aims at and must be measured up to. In order to assess the quality of teaching and learning in higher education, many scholars advocate the consideration of diverse learning outcomes, not just knowledge gains (e.g., Helmke & Schrader, 2010; Shavelson, 2010).

Different than in the previous subchapters on input and process variables, the scientific questions regarding the output dimension of higher education teaching are not so much concerned with determining criteria for quality, which need to be justified by an effect on learning. The desirable learning outcomes in higher education are set on the basis of a normative decision – by researchers, teachers, students themselves, or the society as a whole. Research questions pertaining to the output dimension centre more on the measurement of quality indicators.

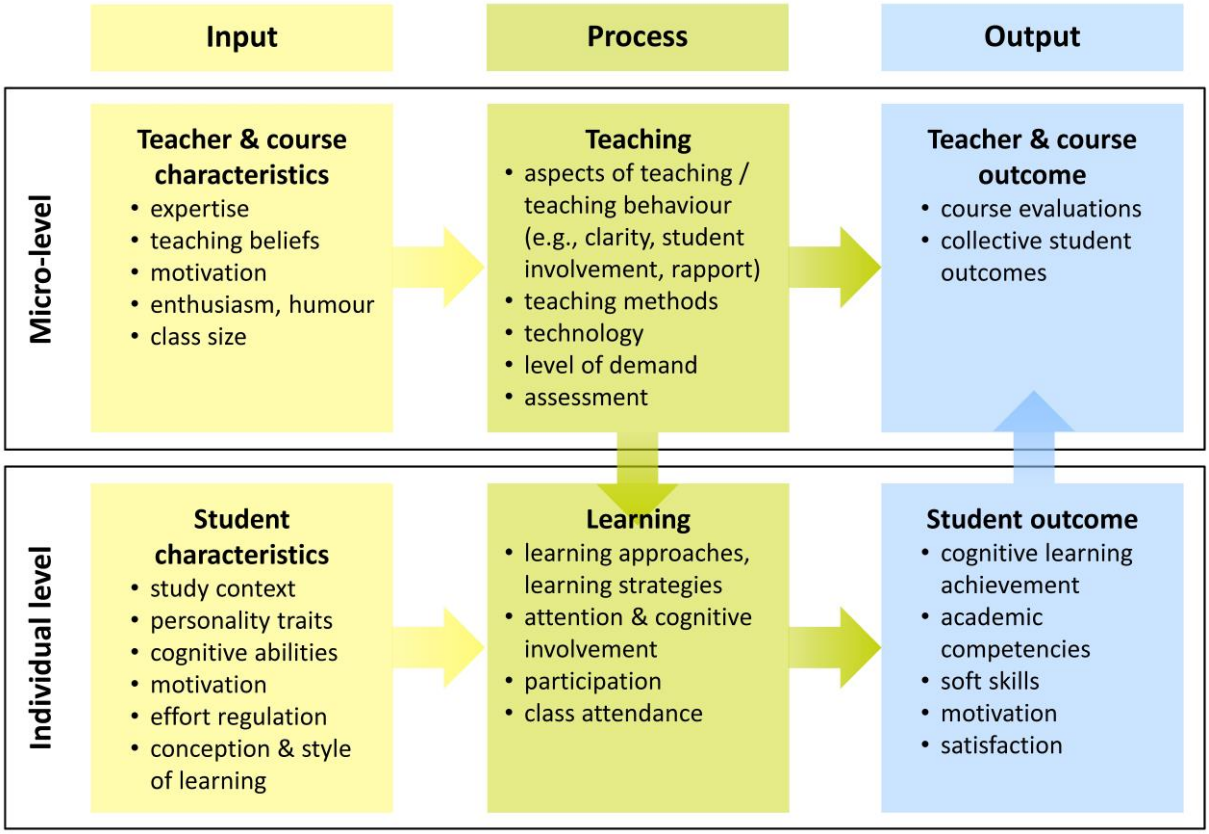

Figure 2.2. Compilation of variables in the input, process, and output dimension of higher education teaching on micro-level and individual level.

2.4 Resulting Ideas for this Study

Each of the six areas of input, process, and output on micro-level and individual level holds a number of unanswered questions. Whereas output criteria of higher education teaching are officially set by public law and usually defined in detail within degree programmes, the identification of input and process variables that indicate quality teaching in its normative sense is less clear. Apart from that, a descriptive perspective on higher education teaching that allows for the exploration of variables and effects without prior judgement seems promising. To progress towards both ends, the relations between the various fields, and in particular effects along the three dimensions of higher education teaching need to be focused on. Potential research questions could be: How do different teacher characteristics influence their way of teaching? How do certain aspects of teaching affect student learning and disparate learning outcomes? How can output variables be measured and to what extent does student assessment of teaching, for example, reflect actual teaching?

This study set out to answer these kinds of questions. It sought to capture the portrayed complexity of higher education teaching and learning including relevant teacher and student characteristics on the input side, the processes of teaching and learning and, of course, output variables on course and student level. To cover the whole process of teaching and learning, the study needed a longitudinal design. To further be able to obtain reliable data and allow for proper analysis of both micro-level and individual level, it also required the usage of multiple data sources.

3 The Research Project – Rationale and Overview

There are multiple ways to research higher education teaching. This chapter explains the approach used in the present study and delineates the resulting methodological layout – its design, the sample, the procedure, and the instruments as well as the fundamental analytic approach. Lastly, it presents three aspects of higher education teaching that were investigated in depth.

3.1 Design

Striving to explore teaching as it commonly occurs in today's higher education, a field study with a correlational design was conducted at a mid-size public university in Germany. To allow for thorough investigation of links between input, process, and output aspects of teaching on course and student level, it had a longitudinal set-up and used multiple data sources. Information on input aspects was contributed by university teachers and students, information on the teaching process was provided by teachers, trained observers, and students, and information on output variables was supplied by students.

Sketching a research design that would provide insight into the relations between processes and products of effective higher education teaching, Abrami, d'Apollonia, and Rosenfield (2007) listed some requirements: Most importantly, a study should be constructed in a way that "allows one to assess (…) what teachers do (process) and the impact teachers have on students (product). In particular, the design must control for plausible rival explanations to the causal effects of instructors" (p. 405). As there are numerous variables influencing learning processes, it is difficult to clearly attribute an outcome to a single origin and dismiss all the other potentially effective variables as causes. Ignoring alternative explanations and additional relevant variables, however, may inflict misleading results and jeopardize internal validity. Aside from concerns regarding internal validity, Abrami and colleagues further urged to pay heed to the transferability of findings with respect to different students, teachers, courses, and institutions, that is, to aspire to a high level of external validity and generality as well.

Devising the design for this study, these recommendations concerning internal and external validity were considered. To ensure a high external validity, a naturalistic design was chosen that investigated regular teaching at an ordinary university under real-life circumstances with students following their usual degree programme. As we were particularly interested in effects occurring in higher education teaching across distinct disciplines, courses from diverse disciplines were included. While a field study holds advantages with regard to external validity and while measurements over departments may increase generalizability, such a design comes with a number of challenges pertaining to internal validity: Examining different courses in various areas of higher education, a multitude of aspects varies between the single learning arrangements. Not only do the teachers differ in experience, expertise, temper, and personal attitude towards teaching; the timing, size, and format of the course, the relevance and level of difficulty of the learning content, and the group of participating students with their individual preconditions vary as well. Hence, comparing learning in different courses, it may be hard to identify the aspect that brought about a certain process or product.

To limit variability and increase comparability between courses, it was decided ex ante to focus on the two most common course formats in German higher education, which are also found in many disciplines: lectures and seminars (cf. Chapter 3.2). Both formats imply regular teaching sessions (in contrast to a lot of self-regulated learning time out of class, e.g., in project courses) and the responsibility of a faculty member (in contrast to senior students, e.g., in repetitive or practical course formats).

To be able to deal with the remaining diversity in teaching settings, particular effort was made to consider a number of variables that may have an impact on learning in higher education aside from the aspects of primary interest. Examples are the subject discipline, course size, or students' preconditions. Even though the inclusion of these variables cannot guarantee internal validity, it might help to eliminate some rival explanations, and, on the basis of sound theoretical consideration, might increase the probability of a realistic representation of the learning situation and a more reasonable estimation of effects.

Lastly, an important limitation that comes with the inclusion of a wide range of courses – both subject-wise and level-wise – shall be emphasised here: Unlike in laboratory settings or multisection studies – where the effects of specific instructional aspects or teachers with varying characteristics are evaluated under constant conditions, e.g., in identical courses – amount, level, and subject matter of the learning contents are utterly incomparable. Therefore, it was not possible to deploy a final test or another measure of knowledge gain that would deliver equally valid results over all the courses (cf. Chapter 2.3.3). As grades at most present valid indication for quantity or quality of the knowledge gained within a single course and hardly denote the learning achievement in one course to be higher than in another, they unfortunately do not pose a reasonable measurement of outcome when input and process aspects of different courses are examined. This is not only due to large deviations between grading styles of single teachers and possibly diverging grading cultures in distinct disciplines (cf. German Council of Science and Humanities, 2012). It is inherent in the lack of comparability of the learning content studied, the level of difficulty, the respective quantity, the depth of understanding, and, most importantly, the different kinds of possible outcomes aspired to by the teacher (e.g., change of attitude or development of reflection ability vs. construction of basic knowledge). The only way to tap learning achievement and obtain meaningful information about the outcome of teaching in diverse higher education settings is information from the students' side. Thus, in this study the assessment of personal interest, of the development of specific competencies, and of the knowledge attained in a particular course via self-report was accepted as a trade-off for inspecting disparate courses in various disciplines.

3.2 Sample

The study was conducted at a public institution of higher education, the University of Kassel, Germany. With about 25,000 students and 3,300 staff members it can be considered a mid-size university (see www.mittelgrosse-universitaeten.de). It offers a broad range of subjects, including humanities, social sciences, economics, and mathematics and natural sciences among others. According to the Europe Teaching Rankings conducted by the Times Higher Education (2018), the

University of Kassel is placed in the lower middle of the 242 institutions that were compared with regard to their teaching quality, and ranks 19th out of 31 German universities included.[3] The score is made up of several key figures, such as the staff-to-student ratio or the graduation rate, but also contains a reputational measure provided by scholars working all over Europe and student reports on interaction with teachers, opportunity for collaborative learning, access to learning materials, the quality of their learning environment as well as on aspects directly linked to their learning processes such as the support of critical thinking, reflection of learning content, application to real-world problems, and a sense of challenge.

The teachers included in this study were selected considering a number of parameters: First, they were required to have a doctoral degree to make a minimum of teaching experience and a certain self-awareness and stability as a university teacher more probable. For newly teaching faculty taking part in a study on teaching quality would certainly have meant considerable additional pressure; apart from that, a certain amount of teaching experience is necessary to form personal views and convictions about teaching, which were also investigated within the study. Second, if possible the participants were supposed to both teach and do research to be able to weigh and compare the two tasks with regard to personal value, perceived competence, or allocated time. In Germany, most university faculty have duties in research and teaching alike; there are only few fulltime teaching posts. Professors and research assistants – if not engaged in independently financed projects – usually split their time between the two tasks.

As it was the explicit aim of this study to investigate effects of higher education teaching that are valid across disciplines, the author strove for a balanced sampling that represented a wide variety of subject disciplines and reflected the professional status distribution in the selected departments. The rather laborious recruitment processes with at times complicated and protracted decision-making on the part of the teachers made it quite difficult to attain the targeted proportions, though. Aside from this endeavour, it was not possible to secure a randomized sample that would have been representative with regard to teaching quality because participation, of course, was voluntary. Therefore, it is very well possible, despite all the effort to convince reluctant candidates in particular, that the participating teachers present a positive selection. Due to the lack of available and meaningful data on the teaching quality of single teachers, there was no way of consciously avoiding a positive selection and working towards a representative sample with regard to teaching quality. Likewise, there is no way now to establish to what degree the final teacher sample is representative in their teaching with any frame of reference – their subject, the university, Germany, or even worldwide. However, whereas this limitation does prevent the production of generalizable data on means or variances of teacher characteristics and aspects of their teaching, it still allows for the detection of relations between different aspects of higher education teaching, which was the aim of this study.

[3] As the higher education institutions included in the ranking are a convenience sample, largely dependent on the availability of data, and thus not representative for the whole of European higher education institutions, the informative value of this ranking is limited. For inclusion criteria and other methodological issues, check: https://www.timeshighereducation.com/world-university-rankings/europe-teaching-rankings-2018-methodology

So what is the result of the recruitment of the participating teachers? How is the teacher sample of this study made up? All in all, 180 university teachers were asked personally to take part in this study with the incentive of receiving feedback on their teaching. Of these, 87 filled out the online questionnaire and 85 completed the guided interview. The 79 teachers participating with full data sets make up 44% of the faculty contacted initially. Of these university teachers 61 held professor positions (77%), 12 were research assistants (15%), and 6 worked as fulltime lecturers (8%); 58 of the teachers were male (73%), 21 female (27%). The teacher sample comprised 14 persons (18%) aged between 31 and 40 years, 26 (33%) between 41 and 50 years, 33 (42%) between 51 and 60 years, and 6 (7%) between 61 and 65 years; their teaching experience in higher education ranged from 3 to 37 years (M = 18.4, SD = 8.2). The subject disciplines they were affiliated with are depicted in Figure 3.1.

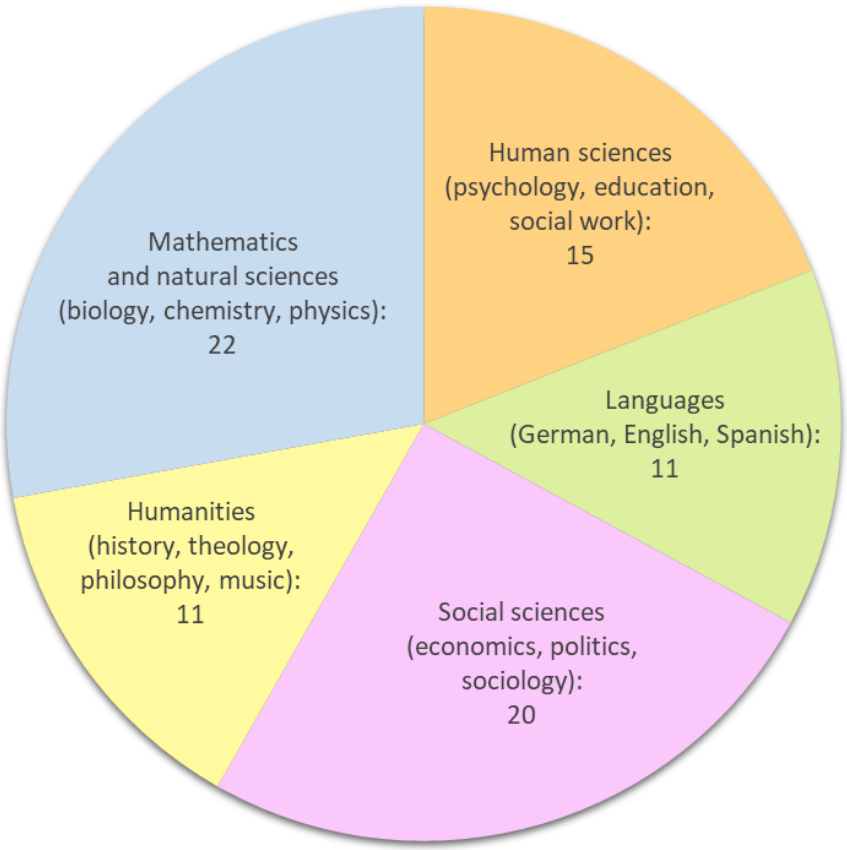

Figure 3.1. Distribution of the participating university teachers over different disciplines.

To increase comparability among the investigated courses (cf. Chapter 3.1) and to ensure the feasibility of the various measurements, two primary conditions were formulated: 1) Only lectures and seminars with regular, mostly weekly sessions, where 2) the teacher would be present and responsible himself were to be selected. As a result, tutorials carried out by senior students, project seminars or online courses with the better part of the learning done outside of class, and laboratory courses, excursions, or block seminars were left out of this study from the outset. The resulting

sample of courses comprised 48 lectures and 32 seminars (one teacher participated twice). They covered a broad range of disciplines; 20 courses could be subsumed in the category of natural sciences and mathematics, 60 were situated in other disciplines.[4] Choosing a course for investigation, the teachers were further asked to pick one in which at least 20 students would be enrolled. This minimum participation was set to ensure a sound analysis of student data. However, as some teachers in some disciplines hardly ever teach courses of that size, and as most teachers do not know the number of participants before the beginning of the semester, this precondition could not be maintained. The final sample of investigated courses varied considerably in size with a minimum of 8, a maximum of 386, and an average of *M* = 59.8 (*SD* = 62.8) students present at the beginning of the semester.

Thirteen teachers had not given that particular course before, other teachers reported to previously having conducted it up to 25 times (*M* = 5.4, *SD* = 5.5). According to the teachers, in 24 courses the learning content did not have any overlap with their field of research; for 23 courses they indicated little overlap and 33 courses were clearly overlapping with own research.

The students enrolled in the courses that the teachers selected for this study amount to a total of 5,765 students. Over all the courses, 4,829 students took part in the entry survey and 2,584 students in the final survey. Participation in the study was voluntary, but students were given the incentive of taking part in a raffle with a tablet computer and ten 20-Euro vouchers as prizes if they completed both surveys. The dropout can mainly be explained by the decreasing attendance over the course of a semester; three courses withdrew after the first survey because of serious sickness of the teachers or too little student enrolment. Overall, 1,716 students were identified by their personal code as the same person taking part in the entry and the final survey.

The students that completed the entry survey were on average *M* = 23.2 (*SD* = 4.3) years old and had been studying for *M* = 4.1 (*SD* = 2.6) semesters, 56.7% of them were female. Of these students 75.5% had the general qualification for university entrance, 22.2% had a specialised qualification, and 2.2% had other qualifications; their high school diploma grade was *M* = 2.4 (*SD* = 0.6) on average on a scale from 1 to 6 with lower numbers indicating higher achievement. Concerning prior knowledge of their respective teacher, 1,560 (32.9%) students reported to having visited another course with the same teacher before, 1,293 (22.4%) students had only heard about the teacher previously, and 1,891 (32.8%) students did not have any information about the teacher before.

[4] Courses by mathematics and natural science teachers that covered the didactics of those subjects were not included in that category; courses in other disciplines such as economics that solely dealt with mathematics, in contrast, were considered as natural sciences and mathematics courses.

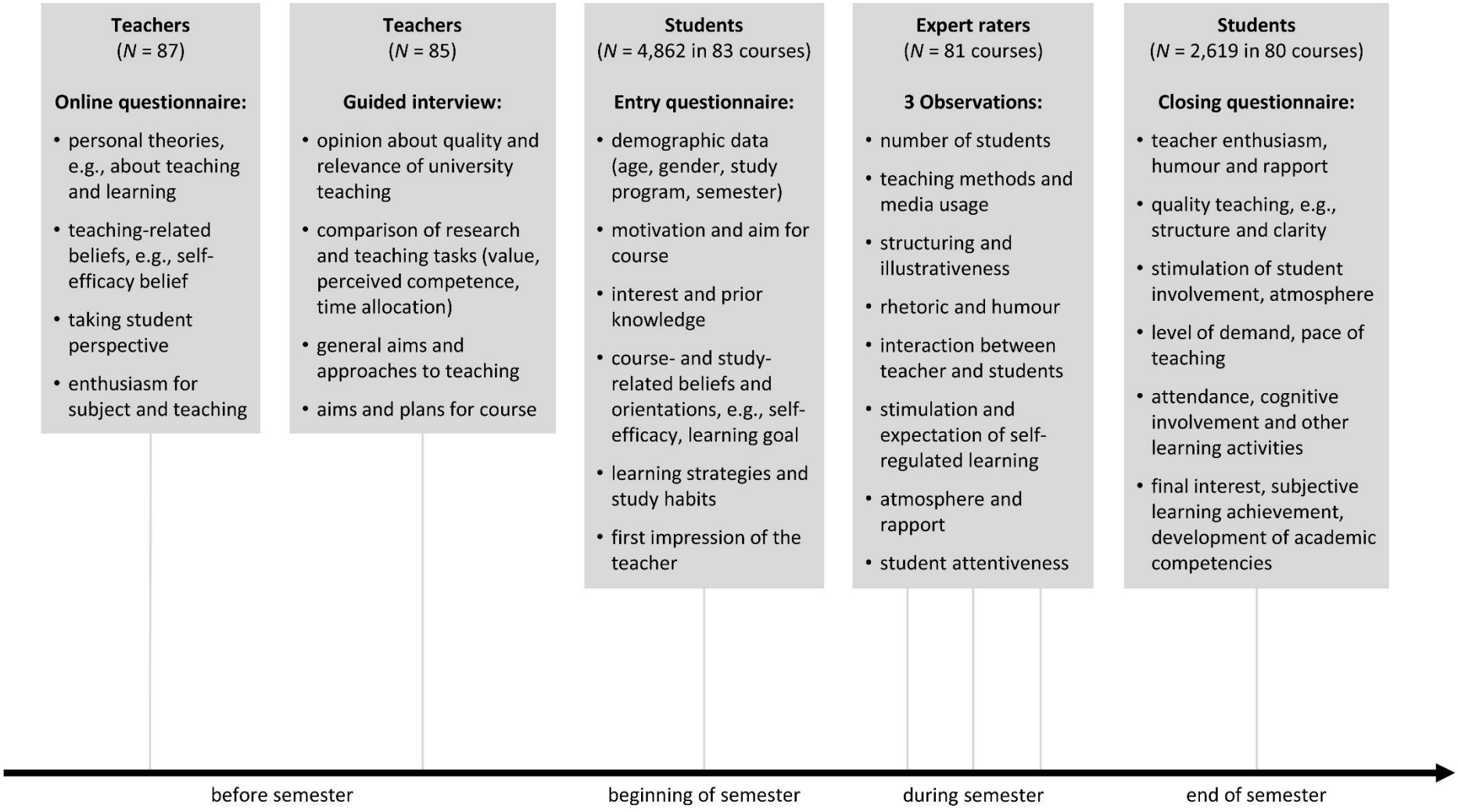

Figure 3.2. Overview of measurements during one semester with data source, total sample size, and exemplary aspects of inquiry.

3.3 Procedure

The study was conducted over the course of three semesters, from the summer of 2014 to the summer of 2015. Each measurement cycle commenced with the recruitment of the teachers, a process, which started about two months before the beginning of the semester. After assenting to take part in the study, the teachers first filled out an online questionnaire and met with the author for a guided interview. Here, personal perspectives on higher education teaching as well as aspired aims and utilized methods were inquired, and the implementation of the student surveys in the selected course was planned. To collect some basic information on the participating students and, in particular, to assess their preconditions, an entry survey was conducted in the beginning of the semester, that is, in the first two weeks, preferably during the very first session.

The observatory part of the study was organised similar to Murray's (2007) research. During the semester two to three trained raters visited each course three times; if possible, one of the three visits was done jointly to be able to determine interrater reliability.[5] They observed the teaching and rated a number of aspects on a standardized rating form. In the very end of the semester, usually during the last session, the students filled out the closing questionnaire retrospectively describing the teaching experienced and providing information on their own engagement within the course as well as on the perceived learning outcome. The various measurements are illustrated in Figure 3.2.

3.4 Instruments

A rough overview of the constructs assessed in this study at the various measurement points by means of questionnaires, interviews, and observations is given in Figure 3.2. Detailed information on all the variables measured with the distinct instruments is compiled in a separate scale documentation (Fischer, 2018).

A core element of this study was the rating of higher education teaching by trained observers. The rating form used to assess various aspects of teaching in a standardised manner was specifically developed for this research project and resembles the instrument of Wibbecke, Wibbecke, Kahmann, and Kadmon (2016). As scholars might be particularly interested in this tool, the rating form was revised and modified on the basis of the experiences and information gathered in this study. To make it available for use and further development to higher education researchers worldwide, it was also translated into English and can be found in the appendix of this thesis.

3.5 Analysis

A central aspect that concerned the analyses conducted within this research project was the multi-level structure of the data. The students were nested within courses and thus systematically shared part of their variance. This fact needs to be considered in any analyses involving student data.

[5] The raters were senior student research assistants who were employed shortly before or after finishing their degree and, thus, had a lot of experience in higher education teaching. They completed a training of about 24 hours, which involved the meticulous study of the rating instrument, the assessment of seven online lectures, joint live observations in four disparate university courses, and extensive discussions about the instrument and single ratings with the author.

Methods that disregard grouping build on the assumption of independence of observations, and may hence lead to erroneous conclusions. In particular, standard errors may be underestimated as the effective sample size is being overestimated. In consequence, statistical inference estimations may be incorrect, indicating effects as significant even though they are not (type I error), and the precision suggested by the confidence intervals may be overrated (Geiser, 2011, pp. 199-200).

Apart from the statistical considerations pertaining to the adequate handling of clustered student data, the question of how to deal with variables measured on two distinct levels also arises, when student data are to be combined with teacher or course variables, which is the case for almost any question pertaining to the effectiveness of teaching. Of course, there are different ways to acknowledge the nested data structure and to relate information on student level to variables on teacher and course level. The choice of the procedure, naturally, depends on the research question and the resulting focus of analysis. Three possible approaches are sketched here briefly:

1) A conventional way to deal with data from distinct levels is to aggregate the observations from the lower level and conduct the analysis on the upper level only. This can be problematic – especially when these upper-level effects are unduly interpreted on the lower level (cf. Hox, 2010; Lüdtke et al., 2008; Snijders & Bosker, 2012). Only under special circumstances may this procedure be regarded as appropriate: Obviously, the researcher must be interested in effects on the group level only; as soon as processes on the individual level became relevant as well, aggregation would be inacceptable. Lüdtke et al. (2008) explain that especially in the case of formative variables – variables measured on the lower level that "form" a group indicator which has a different meaning (e.g., gender and gender ratio, or socio-economic background) – aggregation can be adequate as the mean displays a true characteristic of the group, at least with good sampling. However, as aggregation ignores the varying reliability of group means, which is influenced by group size and the intraclass correlation (ICC) of the variable, and thus presumes perfect measurement, necessary preconditions to use this procedure for reflective variables are sufficiently large groups and high ICCs (cf. Lüdtke et al., 2008; Snijders & Bosker, 2012, p. 14). This approach was applied in the first substudy (Chapter 4), which investigated relations between teacher characteristics and their way of teaching.

2) The second approach of multi-level modelling has been evolving over the past two decades. The advantage of multi-level modelling over conventional ways of analysis such as aggregation lies in the utilisation of not only the group mean but also its reliability by considering the variance of the individual ratings plus the group size. The values on the group level are weighed according to their precision. This is particularly relevant when relations between variables are estimated, as reliabilities of the measurements influence the covariance (Nezlek, Schröder-Abé, & Schütz, 2006). Thus, when estimating contextual effects, the course means are treated like a latent variable, which takes into account their measurement error. This is particularly advantageous for the merging of reflective variables – variables that are measured by individuals but directly "reflect" a generic group level construct, which in turn is assumed to cause the observations on the lower level. Here, within-group variation represents a lack of agreement among individuals. As multi-level analyses infer latent unobserved group means taking into account the imperfect reliability of the observed group mean, which depends on the number of units within the group and the ICC, the

estimates of group-level coefficients are corrected for unreliable measurement (Lüdtke et al., 2008). This approach was employed in the second substudy (Chapter 5), which examined the effects of certain teaching methods on student learning.

3) Most recent advancements of multi-level modelling recommend to factually decompose individual data into its two variance components by introducing latent factors on individual and group level into the model (e.g., Lüdtke, Marsh, Robitzsch, & Trautwein, 2011; Preacher, Zhang, & Zyphur, 2016). This procedure allows for explicit working with the distinct parts of variance, which might become necessary, for instance, when conducting moderation analyses, in which a specific variance part of a variable is hypothesized to cause an effect (cf. Preacher et al., 2016). In addition to the latent modelling of the two levels inherent in individual data stemming from different groups, some scholars suggested to also model the variables latently by including the factor model, where a factor represents the shared variance of single items measuring a construct, to better account for measurement error due to the used instrument. This rather complex but possibly particularly precise way of analysing multi-level data is labelled doubly latent modelling (Lüdtke et al., 2011). The doubly latent analyses, simultaneously correcting for measurement error and sampling error, render unbiased estimates on group level under favourable conditions. However, approaches considering only one of the two sources of error can outperform them and may provide more accurate estimates if the ICC, the number of groups, and the number of persons within the groups are low (Lüdtke et al., 2011). As the latent modelling of individual data on two levels provides the possibility to differentially tackle specific shares of variance, it was applied in the third substudy (Chapter 6), which investigated potential impediments of the assessment of higher education teaching by students.

Recommendations on the sample size for the analysis of multi-level data started off with at least 30 groups with 30 individuals each as a rule of thumb (Kreft, 1996, as cited in Hox, 2010, p. 235), but were adjusted towards a prioritisation of groups over individuals if the main interest is in group-level effects: Hox (2010) proposed 50 groups with 20 individuals each if cross-level effects are being investigated. With a growing interest in the variation between groups, however, the number of groups should further increase while the number of individuals is of less importance. Thus, keeping the total sample size constant, 100 groups of 10 people each should be favoured (Hox, 2010).

3.6 Preview of the Substudies

This study sought to capture the complexity of teaching and learning in higher education. The longitudinal design and the numerous student, teacher, and course variables pertaining to the input, process, and output dimension of higher education allow for the investigation of different questions with regard to higher education teaching. To underline the wide range of issues that arise when talking about the quality of higher education teaching, three disparate aspects were chosen for further investigation. Each substudy focused on another dimension of teaching in higher education. All three studies followed up on long-standing strands of higher education research (cf. Dunkin & Barnes, 1986; McKeachie, 1990) and tried to advance the respective fields.

The first substudy considered the impact of factors of the input dimension on aspects of the process dimension, that is, teacher characteristics and their effect on teaching behaviour. Picking up previous work on teaching beliefs, the research on the significance of teacher characteristics was expanded by the introduction of another, potentially relevant variable: teachers' task value of teaching. The variables' respective effects on various aspects of teaching were examined.

The second substudy focussed on the process dimension of higher education and took up the old question of the effectiveness of distinct teaching methods. The study expanded the existing literature by inspecting natural teaching assessed by expert observation and by investigating not only the influence on the students' immediate learning processes but also their learning outcomes. To meet the claim of Shavelson (2010) and to be able to detect differential effects, the students' final interest and the development of competencies were used as outcome measures aside from the gain of subject knowledge.

The third substudy was devoted to the outcome dimension and had a slightly different take on the topic of quality in higher education teaching. Instead of exploring process effects of higher education, it was concerned with the measurement of teaching quality. The investigation was situated in the long tradition of research on student evaluation of teaching. As scepticism with regard to the validity of teaching assessment by students is still prevalent, and a number of recent studies displayed methodological deficiencies, this substudy contributed to the debate by using multiple data sources and applying state-of-the-art statistical analyses.

Figure 3.3 provides an overview of the three substudies, which are subsequently described in the following chapters. As they were designed as individual publications, each of the substudies begins with a separate introduction providing an overview of the specific field of research and deducting the respective research question.

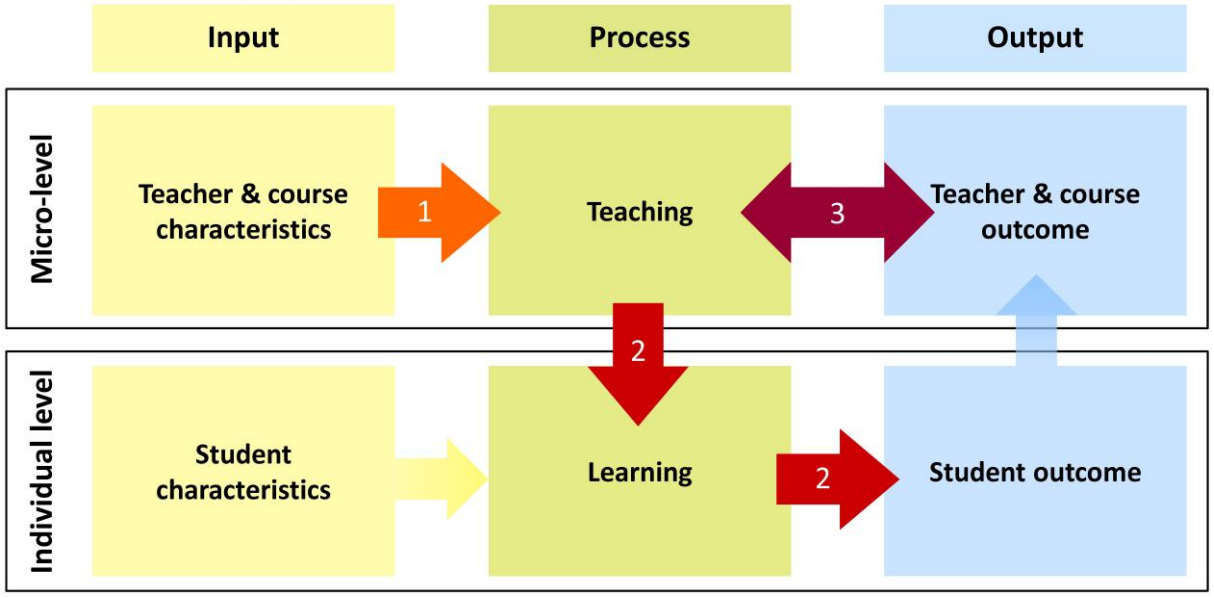

Figure 3.3. Simplified illustration of the three substudies within the framework of higher education teaching.

4 Substudy on the Input Dimension of Higher Education Teaching [6]

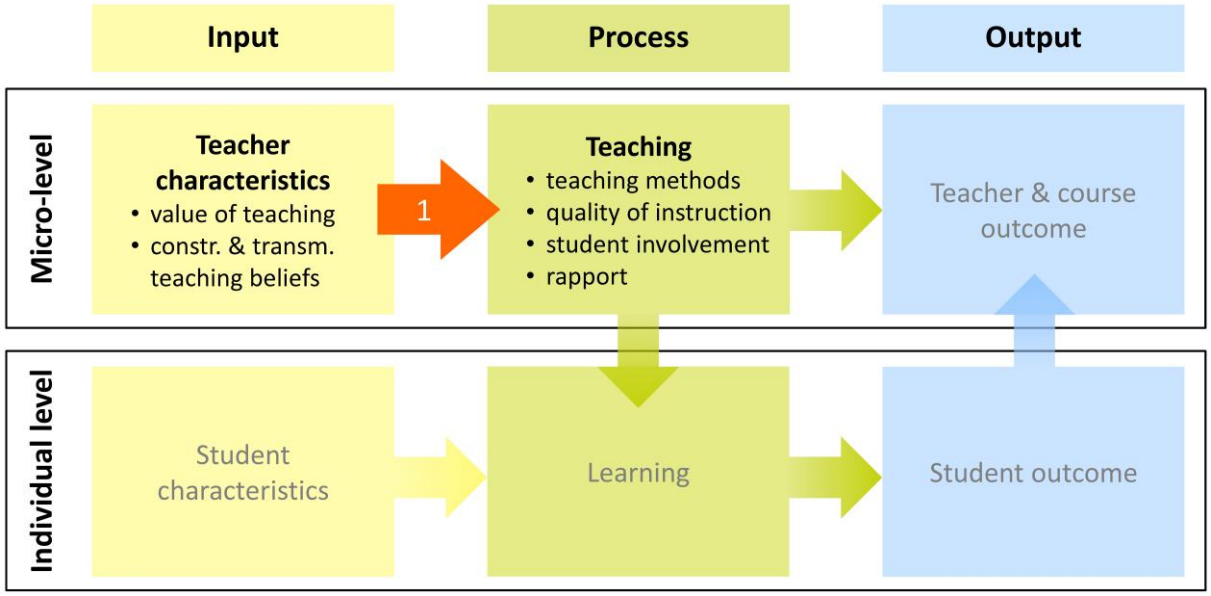

Figure 4.0. Illustration of the first substudy within the framework of higher education teaching.

[6] This chapter was co-authored by Martin Hänze and published in Educational Psychology. Please cite: Fischer, E., & Hänze, M. (in press). How do university teachers' values and beliefs affect their teaching? *Educational Psychology*. doi.org/10.1080/01443410.2019.1675867 (open access)
Minor modifications (e.g., numbering of tables and headings) were made here to increase consistency within the dissertation. Sources are cited in a way that also allows for separate reading (citation of all authors with first mention in chapter).

How Do University Teachers' Values and Beliefs Affect Their Teaching?

Abstract

This study investigated university teachers' characteristics and their influence on teaching practices: Coming from expectancy-value theory, teachers' personal value of teaching was introduced as a possibly relevant variable and examined along with constructivist and transmissive teaching beliefs as to how they affect various aspects of university teaching. The sample consisted of 79 university teachers, whose data were combined with the teaching assessment of 2,552 students enrolled in their courses and of external observers. Multiple regression analyses showed that value of teaching affected observed structuring, student active involvement, and rapport. Transmissive beliefs impacted the observed quality of instruction, and constructivist beliefs were positively related to student-reported clarity of instruction, the stimulation of student involvement, and rapport. Notably, the predictors displayed a data source specific result pattern. As potential reasons for the link between student-reported variables and constructivist beliefs a general factor in the students' assessment and a survivor bias associated with teachers' constructivist beliefs are discussed.

Keywords: expectancy-value theory; higher education; task value; teaching; teaching beliefs

4.1 Introduction

The significance of university teachers and their teaching for student learning in higher education is uncontested. However, the riddles of how quality teaching may be achieved and which teacher characteristics might be beneficial are still largely unresolved. Looking to answer these questions, different teacher characteristics have been investigated – among others enthusiasm (e.g., Kim & Schallert, 2014; Patrick, Hisley, & Kempler, 2000), goals (e.g., Daumiller, Grassinger, Dickhäuser, & Dresel, 2016; Wosnitza, Helker, & Lohbeck, 2014), and goal orientations (e.g., Han, Yin, & Wang, 2016). Special attention has been paid to teachers' personal beliefs about teaching – also referred to as subjective theories, teaching conceptions or orientations (e.g., Kane, Sandretto, & Heath, 2002; Kember, 1997; Mesa, Celis, & Lande, 2014; Norton, Richardson, Hartley, Newstead, & Mayes, 2005; Trigwell, Prosser, & Waterhouse, 1999). Diverging beliefs about suitable approaches and aims of university teaching are assumed to result in equally diverging ways of teaching and thus to influence the students' approaches to learning. Notwithstanding the numerous studies exploring teachers' beliefs and their impact, it remains rather unclear, how exactly distinct beliefs affect university teaching. So, along with Kane et al. (2002) we are still "unconvinced […] that the relationship between teachers' espoused beliefs and their teaching practice has been investigated sufficiently thoroughly to draw any definite conclusions" (p. 204).

Aside from deficits in the existing research on university teachers' characteristics, we think that another potentially fruitful variable might not even have been considered yet – the personal value attached to the teaching task. A high task value has been shown to increase corresponding effort and performance (e.g., Cole, Bergin, & Whittaker, 2008; Dietrich, Viljaranta, Moeller, & Kracke, 2017; Eccles & Wigfield, 2002; Grant, 2008; Turney, 1974). Given the extensive research based on expectancy-value models in both educational and work-related contexts, it is surprising that the fundamental motivational component of task value has not yet been considered with regard to higher education teaching.

This paper introduces the value construct into the field of higher education and compares it to the well-established concept of teaching beliefs as to how they affect various aspects of university teaching. We wanted to find out, to what extent teaching is a matter of how seriously it is being taken by the instructors (value), and in what way it is influenced by how they conceive and approach it (beliefs).

4.1.1 Expectancy-Value Theory and Higher Education Teaching

Modern expectancy-value theories are based on Atkinson's (1957) expectancy-value model and link performance, persistence, and choice to individuals' task specific expectancy of success and the value attached to the task and its consequences. Expectancy-related beliefs are subjective appraisals about how well one will do on an upcoming task, the estimated probability of success. They are mingled with ability beliefs, which are individuals' perception of their competence in a given activity, and very similar to Bandura's (1997) construct of self-efficacy. Value-related beliefs concern the incentive power of the pursued goal. As one of the major developments of the model, particularly in the educational field, the ascribed task value has been differentiated into four components: 1) attainment value, implying the personal importance of a given task, 2) intrinsic value,

i.e. the positive emotions felt while doing this task, 3) utility or usefulness, describing how the task at hand fits into one's future plans and serves one's greater goals, and 4) cost, referring to the disadvantages or limitations with regard to other activities, time, effort, or emotional strain that come with doing the task (Eccles & Wigfield, 2002; Wigfield & Eccles, 2000). While the first two components can be characterized as intrinsic in nature, the latter two are more extrinsic motivations.

Expectancy-value theory has been quite fruitful in educational contexts, explaining behaviour and success of learners with distinct motivational resources. For various age groups it was repeatedly shown that students' expectancies of success are closely associated with their subsequent performance, whereas students' subjective task values predict their intentions and choices about future courses or careers (Jones, Paretti, Hein, & Knott, 2010; Perez, Cromley, & Kaplan, 2014; Wigfield & Eccles, 1992; 2000) but also effort and persistence in achievement-related activities (Cole et al., 2008; Gonzáles, Rodríguez, Faílde, & Carrera, 2016; Dietrich et al., 2017). While most studies on student learning investigated both components of the expectancy-value model, research investigating the motivational determinants of teacher behaviour solely focused on expectancy beliefs, mostly using the self-efficacy construct. For school teachers self-efficacy has been shown to be positively related to their teaching performance; teachers with high self-efficacy are specifically better at coping with a range of problematic student behaviours, tend to use more proactive, student-centred classroom practices and more diverse instructional strategies, and enhance student motivation (e.g., Klassen & Tze, 2014; Zee & Koomen, 2016). Studies examining the influence of university teachers' self-efficacy are rare, but seem to point in a similar direction (e.g., Baleghizadeh & Shakouri, 2017). No studies whatsoever were found on the role of the teachers' value beliefs concerning their teaching task. While it may only be limitedly reasonable to inquire the value school teachers place on the task of teaching – as it is their main task and they do not have a lot of options for own priorities aside from teaching-related work, the situation is quite different in the higher education context. University teachers usually have a range of different tasks at hand, with teaching being only a part of their professional responsibilities, and have to allocate time and effort to distinct tasks according to their priorities. In that respect, they are quite similar to other employees that have certain autonomy and the freedom to prioritize tasks according to their values. Therefore, it seems worthwhile to also take notice of the research on the motivational power of value in working contexts.

The expectancy-value theory in the field of organizational and industrial psychology is based on the work of Vroom (1964), who expanded Atkinson's (1957) model. There is convincing evidence underlining the importance of value beliefs in the professional world. Grant (2008), for example, was able to demonstrate that perceived task significance boosted job performance: In three experiments with callers and life guards he could repeatedly show that the employees in the condition receiving the task significance intervention increased their job performance in comparison to the ones in the control condition and their own previous performance. Sun, Vancouver, and Weinhardt (2014) explored the specific functions of expectancy and value beliefs during the stages of goal realization. The authors found that both motivational components jointly predicted goal choice, but played independent and opposite roles in goal planning processes, which determine the amount of resources allocated to achieve the goal: Here, the expectancy negatively affected the allocation of

resources, whereas value retained its positive effect. The value component was further refined by Mitchell and Albright (1972) who specified intrinsic rewards related to job content as feelings of self-esteem, self-fulfilment, and worthwhile accomplishments, whereas extrinsic rewards included factors like a promotion, the salary, and job security. Turney (1974) specifically investigated the intrinsic value component next to the values of job outcomes as well as the expectancy that one's own performance will be successful. The intrinsic activity value, as he called it, was a much stronger predictor of motivation, effort, and performance than the joint term of expectancy and extrinsic value. Turney concluded that, in particular, low-structured professional environments "may provide extensive opportunities for […] employees to perform activities which they find intrinsically rewarding" (p. 78). Thus, for intrinsic motivation and value to become important, the work context is highly relevant – if employees have choices and can set their own priorities, these motivational aspects may deploy their power.

Examining the research on the effects of value and expectancy beliefs, the significance of the measurement of the criteria becomes apparent. First, the choice of the construct is decisive: In their meta-analysis van Eerde and Thierry (1996) found that attitudinal criterion variables like preference and intention show stronger relations than behavioural ones like choice, effort, and performance; the relations grow weaker the further they are located within the course of action. Second, the source of information is critical: Various studies present stronger relations for self-report measures of criteria like effort or performance than for ratings by external evaluators (Pritchard & Sanders, 1973; Schwab, Olian-Gottlieb, & Heneman, 1979; Turney, 1974).

As mentioned earlier, there is no research to be found that investigates university teachers' personal value of the teaching task with regard to its effect on teaching. Values of higher education teachers have only been of interest with respect to the prominent rivalry between the tasks of teaching and research most university teachers face. Neumann (1996) detected three approaches to examining the teaching-research nexus in higher education literature: 1) personal views and commentaries, usually qualitative, 2) correlations between measures of teaching effectivity, mostly assessed via student evaluations, and research productivity, i.e. number of publications etc., and 3) surveys of academics' work preferences and time allocation regarding teaching and research. A meta-analysis on the second approach, the relationship between the quality of teaching and research, by Hattie and Marsh (1996) suggested that the two are essentially unrelated. A recent study by Cretchley et al. (2014) illustrated strikingly how university teachers tend to appraise teaching and research tasks. Cretchley and colleagues investigated university teachers' priorities, beliefs, and behaviours with Australian academics in the fields of natural sciences, information technology, and engineering. The participants ranked 16 research-related activities and 16 teaching-related activities according to the importance they had for their job satisfaction, their role model behaviour, and their professional career. The findings revealed unequivocally a clear favouring of the research-related activities in all three value frames and were remarkably consistent over universities, academic levels, age groups, and gender.

Despite the considerable research pressure and the range of administrative duties that usually come with university teachers' positions, they can still be characterised as comparably low-structured professional environments, which give leeway for employees to follow their priorities.

Therefore, we assume that university teachers will try to allocate time and effort according to their personal task values and that the corresponding performance, be it with regard to research, administration, or teaching, may vary. A higher value of teaching may, for instance, lead to more thorough preparation, which may result in clearer structuring, better thought-out questions, and, possibly, in more student activation.

4.1.2 Teaching Beliefs in Higher Education

In contrast to teachers' personal values their beliefs about teaching are a well-researched topic in higher education. To describe what university teachers think about teaching and how they understand their role in the learning process, various terms have been used: *beliefs, conceptions, personal practical knowledge, orientations to teaching, subjective theories,* and *attitudes* (cf. Kember, 1997; Trigwell & Prosser, 1996). The initial research phase in the 1990s was characterized by a number of qualitative studies that explored teachers' thoughts and convictions about teaching and tried to deduce categories and schemas. Samuelowicz and Bain (1992) summarized the first explorative studies on university teachers' conceptions of teaching and reported broad agreement about a conceptual range from teaching being primarily conceived as information presentation to teaching as facilitating student learning. They proposed five conceptions of teaching that differ with regard to knowledge acquisition, students' preconceptions, the expected outcome of learning, the directionality of teaching, and control of the content. Kember and Gow (1994) used the terms of *knowledge transmission* and *learning facilitation* to describe two distinct orientations to teaching. According to their model each of these broad orientations comprises several specific facets: Learning facilitation includes problem solving, which involves independent learning and critical thinking, an interactive teaching approach, the understanding of teaching as a facilitative procedure, interest in student concerns, and student motivation. Knowledge transmission, on the other hand, is associated with training for a specific job, use of media, knowledge of the subject as a prime requirement for faculty, and imparting information, i.e. transferring knowledge by presenting it as clearly and accurately as possible. In an attempt to synthesize the early exploratory work, Kember (1997) reviewed 13 qualitative studies and summarized their findings in two superior orientations: *teacher-centred/content-oriented* and *student-centred/learning-oriented*. The former emphasizes the communication of content, while the latter adopts a more developmental approach and focuses on student learning. On the basis of preliminary interviews with natural sciences lecturers, Trigwell and Prosser (1996) developed an instrument to measure the intentions and strategies for teaching in higher education, the *Approaches to Teaching Inventory* (ATI). The intentions denote the aims of the teachers and are defined by the two poles of transmitting the content of the subject to the students and helping students change their conceptions of the content. The strategies refer to the way of teaching and range from methods with a focus on the teacher to methods with a focus on the students. Correlational analyses showed an association of the aim of information transmission with teacher-focused strategies, and the aim of conceptual change with student-teacher interaction and student-focused teaching strategies. Combining intentions with strategies, Trigwell and Prosser suggested five ordered approaches to teaching, labelling one extreme *teacher-focused strategy with*

the intention of transmitting information to students and the other *student-focused strategy aimed at students changing their conceptions*.

With a growing body of research, it became apparent that teachers might have what Samuelowicz and Bain (1992) termed *ideal* and *working* conceptions of teaching. Kane et al. (2002) underlined the importance of distinguishing thoroughly between, as they called them, *espoused theories* of action and *theories-in-use*. Whereas espoused theories encompass personal aims and intentions and will be relayed by most people if asked about their behaviour, theories-in-use actually determine actions, but are predominantly tacit knowledge and thus not articulated easily. Following Kane and colleagues, self-reported beliefs and approaches might therefore well be a close approximation of teaching practice in some cases, but ultimately not a reliable measure.

Higher education researchers were interested in the effects of teaching beliefs on the students' learning already quite early. In one of the first studies investigating the relation between teaching orientation and student learning, Kember and Gow (1994) measured both the instructors' teaching approaches and the students' approaches to learning in 15 departments. Correlations on department level revealed significant negative relations between teacher and student variables: In departments with a stronger teaching orientation towards knowledge transmission, students reported a less deep approach to learning; in departments where teaching was understood more as facilitation of learning, students were much less likely to report the use of surface approaches to learning. A couple of years later, Trigwell et al. (1999) also tested associations of the teachers' approaches to teaching with their students' approaches to learning. Analyses with 48 first year courses in natural sciences showed that the teachers' information transmission/teacher-focused approach to teaching went along with more surface and less deep approaches to learning, whereas students whose teachers adopted a conceptual change/student-focused approach were less likely to show a surface approach. However, as both teachers and students were asked to relate their answers to the same lecture topic, this conformity may also simply depict the learning opportunities in a particular course session.

As the studies that tried to link the students' learning approaches to the teacher's teaching approaches failed to explain how exactly the teachers might impact their students' learning, subsequent studies sought to establish the effect of the teachers' reported beliefs on the actual teaching practice. However, from our point of view, the results are not quite satisfying yet: Attempts to capture teaching behaviour via self-report by formulating items in relation to a specific teaching situation or inquiring the concrete teaching methods used have rendered questionable results. Whenever a construct referring to actual teaching was assessed in teacher self-report, it was hardly empirically distinguishable from the teaching beliefs (e.g., Norton et al., 2005; Prosser & Trigwell, 2006) creating doubts about the validity of the measurement (Meyer & Eley, 2006). Observational studies, on the other hand, led to quite contradictory results: Martin, Prosser, Trigwell, Ramsden, and Benjamin (2000), for example, tried to link university teachers' thoughts and intentions about what is to be learnt and how it might be taught to actual teaching behaviour by first interviewing university teachers and subsequently observing two of their classes. With the interview explicitly asking about the class examined later on, they found close accordance between the previously stated objectives of learning and the observed approaches to teaching. As a limitation, it has to be noted, though, that

the observation was not standardized, but explicitly aimed at finding indications regarding the hypothesis previously developed on the basis of the interviews. Mesa et al. (2014) conducted a qualitative study with college teachers in the mathematics department. They reported an association between the approaches to teaching stated in the interviews and the observed framing talk used to organize the instruction in class, but no relation to the kind of questions asked. Assen, Meijers, Otting, and Poell (2016) studied teacher beliefs and behaviour in a problem-based learning environment. While the teachers largely reported learner-oriented beliefs, the observations showed a more traditional teaching practice. Similar findings were presented by Ebert-May et al. (2011), who investigated teaching beliefs and practice in the aftermath of professional development workshops. The self-reported survey data indicated that the professional development activities had led to significant gains in both knowledge and use of student-centred teaching, but the observation of the participants' classrooms after completion of the workshops revealed a predominant usage of lecture-based, teacher-centred pedagogy. In the absence of prior observations it remains unclear, though, whether the observed low levels of student activation may already have been an improvement. So, all in all, no clear picture evolves in regard to the actual impact of university teachers' beliefs on their teaching practice.

There are notably few studies deriving specific hypotheses from the two teaching beliefs about concrete teaching characteristics. Having delineated the two opposing views university instructors tend to hold about teaching, we propose the following assumptions: The belief that teaching mainly consists in the presentation and transmission of knowledge with the aim of the students accumulating information may, for instance, result in a greater use of teacher-guided methods with, possibly, particularly well-prepared and structured talks. The belief that teaching is essentially a process of supporting the students' learning with the aim of conceptual or even personal change, may, on the other hand, induce a greater use of student-activating methods and other measures to stimulate student involvement, like higher-order questions.

4.1.3 Research Question

This study set out to explore in what way university teaching might be influenced by the value teachers attribute to the teaching task on the one side and by their personal beliefs about teaching on the other side. As stated previously, a number of different relations are conceivable. To capture a wide range of possible effects, four central aspects of teaching, which are commonly considered core elements of university teaching and indicators of quality (Murray, 2007; see also Griffiths, 2009; Morton, 2009), were examined: 1) the teaching methods used, 2) the quality of instruction, i.e. structure and clarity of the teacher's talk, 3) the student involvement, and 4) rapport. The question we wanted to find answers to was: Which effects do teachers' personal task value as well as their constructivist and transmissive teaching beliefs have on university teaching?

As mentioned before, the measurement mode of the criteria has proven to be crucial. In both, studies investigating the effect of teaching beliefs and of task value, self-reported measures of criteria brought about higher correlations than ratings by other evaluators. To ensure a sound measurement and to maximize the informative value of the results, we chose a multimethod approach with three distinct data sources: As teachers, of course, are best able to indicate their

convictions and values themselves, the predictors were assessed via self-report. All four aspects of teaching, however, were observed and rated by trained observers. Apart from that, the students – as recipients of university teaching – provided information as to how they perceived a teacher and his or her teaching as well. Using these two different data sources for the criteria created a more complete picture. That way, it was also possible to further explore the significance of the data source for the respective findings.

4.2 Method

4.2.1 Sample

In preparation of the study, 180 university teachers were asked personally to take part in the study with the incentive of receiving feedback on their teaching; the participation rate was 44%. The resulting teacher sample consisted of 79 lecturers (58 male, 21 female). Of all the university teachers 61 were professors, the other 18 holding research assistant or lecturer positions; as is common in Germany, they had duties in research and teaching alike. On average, their teaching experience amounted to *M* = 18.4 (*SD* = 8.2) years, with a minimum of 3 and a maximum of 37 years.

The investigated courses included both lectures (*n* = 47) and seminars (*n* = 32) and varied considerably in size; the average number of students present in the beginning of the semester ranged at *M* = 82.0 (*SD* = 72.9, *Min* = 9, *Max* = 386) in lectures and at *M* = 25.5 (*SD* = 10.9, *Min* = 8, *Max* = 54) in seminars. In German higher education, lectures and seminars are the two most prevalent course formats. While lectures refer to teachers imparting a broad subject matter – often to large numbers of students, seminars usually focus on more specific learning content and come with smaller learning groups and greater student participation. Thus, the format of the courses both is a proxy for class size and also denotes the way of teaching to be expected in a course. The courses covered a broad range of disciplines, among others philosophy, foreign languages, economics, sociology, and physics. Twenty courses could be subsumed in the category of natural sciences and mathematics, 59 were situated in other disciplines.

Of the students enrolled in the courses, 4,669 students took part in the entry survey and 2,552 students in the final survey. At both measurement points all the students present were asked to take part in the surveys. The drop-out was due to decreasing attendance during the course of the semester, which can be explained by non-compulsory attendance regulations and the choice options in many study programmes leading to course hopping in the beginning of the semester. With respect to their initial interest, the students taking part in both surveys only differed minimally to the ones that were solely present in the beginning of the semester (Cohen's *d* = .11). Participation in the study was voluntary; as an incentive the students were given the option of taking part in a lottery with a tablet computer and ten 20-Euro-vouchers as prizes. The students were on average *M* = 23.2 (*SD* = 4.3) years old and had been studying for *M* = 4.1 (*SD* = 2.6) semesters, 56.1% of them were female.

4.2.2 Procedure

The study was conducted at an ordinary, middle-sized university in Germany in three semesters from 2014 to 2015. The participating teachers were told sketchily that this study explored quality of higher education and would include their personal view as well as student and observer data. Before each semester, the teachers filled out an online questionnaire, were interviewed by the first author, and chose a course for the subsequent investigation. The students enrolled in the respective courses were notified that their teacher took part in a study on university teaching and that their contribution in the two surveys would be crucial. At the beginning of the semester, they filled out an entry questionnaire. During the semester, each course was visited three times by trained observers, who rated the teaching on a standardized form. At the end of the course, the students completed another questionnaire, which assessed their perspective on the course and the teacher.

4.2.3 Instruments

Predictors

To detect the value attributed to teaching by the university teachers, a global measure was chosen. In the guided interview, the university teachers were asked to rank the three main areas of their professional work – research, teaching, and administration – according to the importance they personally ascribed to these areas. Subsequently, a dichotomous variable was computed that indicated whether the teaching task was ranked highest or not – *priority of teaching*. The psychometric properties of this and all the following variables are displayed in Table 4.1.

To assess the teachers' beliefs about teaching, two scales were developed on the basis of existing instruments and the respective research. Although the Approaches to Teaching Inventory (ATI, Prosser & Trigwell, 2006) is widely used, it has been strongly criticized (e.g., Meyer & Eley, 2006). Prosser and Trigwell (2006) themselves suspect item redundancy and advise future administrators to check the applicability of the items and to modify them to reflect the respective context. As the ATI was developed solely with physics and chemistry teachers, we decided to purposefully change the items to pertain to the teaching environments and conceptions of teachers in a broad range of subject disciplines. Redesigning the instrument, our focus lay on creating formulations that captured the theoretical core of the two opposing teaching beliefs. Aside from the applicability in distinct subject disciplines, special care was taken that all the items worked in both lectures and seminars. The two developed scales, labelled *transmissive teaching beliefs* and *constructivist teaching beliefs*, consisted of five items each; both scales showed acceptable internal consistency (see Table 4.1).

The three predictors were not significantly related to one another ($.04 \leq r \leq .19$, $p \geq .09$).

Table 4.1

Psychometric Properties of the Variables

Variable	Source	Measurement format	Cronbach's α[a]	Range[b]	*M* (*SD*)[c]	$ICC_{1,1}/ICC_{1,k}$ for student reports[d]	$ICC_{1,1}$for observer ratings[e]
Predictors:							
Priority of teaching	T_1	Question in guided interview about personal priorities		0 *no*, 1 *yes*	0.5		
Constructivist teaching beliefs	T_2	Scale with 5 items, e.g., "I see my teaching as a support for the students in developing new thoughts and constructing new knowledge."	α = .73	1 *disagree* - 6 *fully agree*	5.3 (0.6)		
Transmissive teaching beliefs	T_2	Scale with 5 items, e.g., "As a university teacher it is my central task to present the learning content well."	α = .68	1 *disagree* - 6 *fully agree*	4.6 (0.8)		
Control variables:							
Course format				0 *lecture*, 1 *seminar*	0.4		
Subject discipline				0 *other,* 1 *sciences*	0.3		
Initial student interest	S_1	Scale with 7 items, e.g., "I find many of the topics covered in this course very interesting."	α_{stud} = .88, α_{course} = .95	1 *disagree* - 6 *fully agree*	4.2 (0.5)	.20/.94	

(table continued)

Criteria:							
Teaching methods:							
Teacher-guided methods	O	Time of teacher-guided methods as a share of total teaching time (in %), e.g., talks, demonstrations		0-100	61.3 (36.0)		.88, .98, .81
Student-activating methods	O	Time of student-activating methods as a share of total teaching time (in %), e.g., reading tasks, exercises, pair- and group-work, games		0-100	5.6 (9.0)		.58, .97, .93
Quality of instruction:							
Verbal structuring	O	Scale with 7 items, e.g., "An outline of the session's learning content is presented and/or the aims of the session are explicated."	α = .73	1 *not at all* - 5 *often or strongly*	2.2 (0.4)		.73, .77, .60
Visual structuring	O	Number of 5 aspects of visual structuring occurring in a session, e.g., highlighting important information		0-5	0.7 (0.8)		.47, .78, .50
Illustrativeness	O	1 item: "The teacher connects learning content with experiences of everyday life and/or uses examples and analogies for illustration."		1 *not at all* - 5 *very often*	2.3 (0.9)		.76, .83, .55
Structure	S_2	Scale with 5 items, e.g., "The course followed a logical structure."	α_{stud} = .86, α_{course} = .91	1 *disagree* - 6 *fully agree*	4.6 (0.4)	.13/.83	
Clarity	S_2	Scale with 2 items, e.g., "The teacher explained the learning content in a comprehensible way."	α_{stud} = .86, α_{course} = .96	1 *disagree* - 6 *fully agree*	4.7 (0.5)	.20/.89	

(table continued)

Student involvement:							
Higher-order questions	O	Number of higher-order questions, e.g., complex, divergent questions			3.1 (4.6)		.90, .80, .80
Student active involvement	O	1 item: "The students are engaged and participate actively."		1 *not at all* - 5 *very much*	2.9 (0.8)		.60, .71, .61
Stimulation of student involvement	S_2	Scale with 5 items, e.g., "The teacher involved the students actively in the course."	α_{stud} = .88, α_{course} = .93	1 *disagree* - 6 *fully agree*	4.6 (0.6)	.32/.94	
Rapport:							
	O	1 item: "The teacher is respectful, friendly and appreciative towards the students. He is attentive, open for other opinions and suggestions, takes student questions and comments seriously and lets them finish."		1 *disagree* - 5 *fully agree*	4.2 (0.5)		.53, .20, .42
	S_2	Scale with 7 items, e.g., "The teacher meets the students with respect."	α_{stud} = .89, α_{course} = .95	1 *disagree* - 6 *fully agree*	5.0 (0.5)	.24/.91	

Note. T_1 guided interview with teachers, T_2 online questionnaire for teachers, S_1 student entry questionnaire, S_2 student final questionnaire, O expert observations. *N* = 79; student data were aggregated course-wise.

[a] Reliability of the student-reported scales is indicated both on student level and on course level. The sample sizes for the analyses on the student-level are for initial interest (S_1) *n* = 4,669 and for the scales assessed with the final questionnaire (S_2) *n* ≥ 2503. [b] Only the scale anchors were labeled. [c] With dichotomous variables the mean score represents the share of the category indicated by 1; the standard deviation is redundant and therefore omitted with dichotomous variables. [d] For student-reported measures the $ICC_{1,1}$ (single measure) was calculated to indicate the agreement of the students concerning their course (shared variance). Additionally, the $ICC_{1,k}$ (average measure) is given to indicate the reliability of the course mean values. [e] To determine the interrater reliability of the observers, the $ICC_{1,1}$ (single measure) was calculated; the three values refer to the three semesters.

Criteria

Striving to capture a broad range of university teaching characteristics, this study included various criteria variables covering four aspects of university teaching: teaching methods, quality of instruction, student involvement, and rapport. The teaching characteristics were assessed by both external observers and the students enrolled in the courses. The observers were advanced graduate students of diverse degree programmes who had completed approximately 24 hours of training and were not enrolled in any of the investigated courses themselves. They inspected the courses (unannounced) three times, using a standardized rating form to describe each session. About 25% of the visits were done jointly to determine the degree of agreement between the raters. Interrater reliability was calculated separately for each semester and was mostly good or excellent (see Table 4.1). The students were asked to assess their course retrospectively in the questionnaire administered in the end of the semester. For student-reported measures the intra-class correlation coefficient $ICC_{1,1}$ indicates the agreement of the students concerning their course (shared variance). Values in the range of .13 to .32 are also found in other studies on course evaluations (cf. Marsh, 2007, p. 333). To check the reliability of the aggregated means, the $ICC_{1,k}$ was calculated and yielded values greater than .80, indicating a high reliability of the student-reported course means. As it was the students' shared perception of the teaching that was relevant, the student data were aggregated course-wise. Generally, we made an effort to assess all four areas of teaching characteristics by both expert observers and students. However, as external observers are better able to concentrate on various specific teaching characteristics, whereas the students focus on the learning content and are busy learning during a course, the single indicator variables vary, and some constructs were exclusively assessed by the observers.

The teaching methods were solely measured by observers. They recorded the amount of time allotted to distinct teaching methods in the course sessions. The respective variables in this study denote the share of teaching time that was used for particular types of teaching methods: The category *teacher-guided methods* comprised teacher talks including demonstrations such as calculations or experiments and the use of music or videos. The category *student-activating methods* contained methods activating all students without direct involvement of the teacher. Examples are reading tasks, exercises, pair- and group-work, or games. Other methods like discussions with the teacher or student presentations, where only a part of the students is directly activated, are considered transitional teaching methods and were not included.

The criterial category quality of instruction comprised five variables that describe the explanations given by the teacher. Three of them were assessed by the observers: verbal structuring, visual structuring, and illustrativeness. *Verbal structuring* was a scale of seven items denoting distinct ways to explicitly structure the learning content, like stating the aim of the session or locating a topic within the discipline. *Visual structuring* indicated the usage of visual aids, like highlighting important information or numbering titles. The variable *illustrativeness* refers to the teacher's verbal explanations and was measured with one global item asking about connections of the learning content with experiences of everyday life and the use of examples and analogies. The two constructs assessed by the students were structure and clarity. *Structure* was measured with a 5-item scale

inquiring whether the students perceived the teaching as well-structured. *Clarity* was assessed with two items on the comprehensibility of the teacher's talks.

Indicators for student involvement were measured by both observer and student ratings. The observers firstly counted the number of *higher-order questions* posed by the teacher during a session. Secondly, they appraised the *student active involvement*, i.e. whether the students seemed engaged and participated actively. The students themselves evaluated retrospectively the teacher's *stimulation of student involvement*. Did the teacher ask questions, did he or she try to involve the students, did he or she make an effort to get the students to think along?

Rapport, i.e. the teacher's friendliness as well as his or her concern and respect for the students, was measured by the two sources quite analogously: Whereas the observers rated the teacher's behaviour globally with one item, the student measure comprised seven items asking, among other, whether the teacher was open for other opinions and interested in the questions and problems of the students.

All the scales measuring the student responses showed good internal consistency – both on student and on course level. For further information on the design of the variables and their psychometric properties, please consult Table 4.1. Table 4.2 informs about the intercorrelations between the criteria. High correlations between the differently measured variables pertaining to the same aspect of teaching support the validity of the respective criterial instruments.

Control Variables

In view of the sample size and common recommendations on the number of variables in regression analyses (cf. Green, 1991), only three control variables were considered. As the course format and the subject discipline inevitably influence teaching practice (cf. Neumann, Parry, & Becher, 2002), these aspects were included in the analyses. Depicting the two kinds of university courses investigated in the study, *course format* was a dichotomous variable (0 *lecture*, 1 *seminar*). Even though the *subject discipline* was assessed quite specifically, the sample size only allowed for a rough distinction between mathematics and natural science courses on the one hand and courses stemming from all other areas including economics, politics, sociology, psychology, education, philosophy, history, languages, and music on the other. Thus, it was also operationalised dichotomously (0 *other,* 1 *sciences*). Apart from these organisational course characteristics, students' retrospective evaluations may be affected by their preliminary motivation. To account for the varying popularity of courses, the average *initial student interest*, measured at the beginning of the course, was considered as a control variable in regressions on student-reported criteria as well.

Table 4.2

Intercorrelations of the Teaching Characteristics (N = 79)

		Teaching methods		Quality of instruction					Student involvement			Rapport	
		1	2	3	4	5	6	7	8	9	10	11	12
Teaching methods:													
1	Teacher-guided methods (O)		-.59**	.13	.62**	.31**	.06	-.02	-.44**	-.71**	-.66**	-.11	-.12
2	Student-activating methods (O)			-.08	-.27*	-.21	-.06	-.14	.17	.45**	.32**	.00	-.04
Quality of instruction:													
3	Verbal structuring (O)				.15	.10	.05	-.03	.23*	.10	-.02	.34**	.12
4	Visual structuring (O)					.00	.12	-.15	-.39**	-.50**	-.42**	-.10	-.20
5	Illustrativeness (O)						-.09	.21	-.07	-.01	-.13	.18	.02
6	Structure (S)							.70**	-.24*	.08	.17	.26*	.55**
7	Clarity (S)								-.12	.15	.28*	.36**	.70**
Student involvement:													
8	Higher-order questions (O)									.56**	.46**	-.04	-.01
9	Student active involvement (O)										.70**	.27*	.23*
10	Stimulation of student involvement (S)											.22	.45**
Rapport:													
11	(O)												.52**
12	(S)												

Note. (O) observer rating, (S) student rating.
* $p < .05$. ** $p < .01$.

Table 4.3

Standardized Regression Coefficients of Predictors on Observed Criteria (N = 79)

	Teaching methods		Quality of instruction			Student involvement		
Predictors	Teacher-guided methods	Student-activating methods	Verbal structuring	Visual structuring	Illustrative-ness	Higher-order questions	Student active involvement	Rapport
Control variables:								
Course format	-.83**	.43**	-.16	-.52**	-.48**	.40**	.46**	-.09
Subject discipline	.11^{+}	-.09	-.08	.22*	-.38**	-.15	-.23*	-.21^{+}
Teacher variables:								
Priority of teaching	-.02	.05	.22^{+}	-.17*	.03	.23*	.19*	.23*
Constructivist teaching beliefs	.01	-.06	-.02	-.10	.01	.05	.08	-.08
Transmissive teaching beliefs	.04	.14	.01	.24**	-.21*	.06	-.02	-.11
R^2	.78**	.24**	.06	.54**	.28**	.34**	.44**	.10

$^{+}$ $p < .10$. * $p < .05$. ** $p < .01$.

4.3 Results

Multiple regressions with the three predictors and the respective control variables –course format, subject discipline, and students' initial interest with student-reported measures – on single criteria variables were conducted. The results are displayed in Table 4.3 and Table 4.4. The analyses yielded the same result pattern when computed as two-level regression analyses in Mplus 7 (Muthén & Muthén, 1998-2014). We decided to present multiple regression analyses on the course level with aggregated student data because the relevant effects concerned the course level only, and one-level regression analyses were the most straight-forward and simple way of analysis.

Table 4.4

Standardized Regression Coefficients of Predictors on Student-reported Criteria (N = 79)

	Quality of instruction		Student involvement	
Predictors	Structure	Clarity	Stimulation of student involvement	Rapport
Control variables:				
Course format	-.11	-.19^{+}	.41**	-.06
Subject discipline	.03	-.25*	-.22*	-.17
Initial student interest	.21^{+}	.42**	.06	.10
Teacher variables:				
Priority of teaching	.16	.14	.13	.18
Constructivist teaching beliefs	.10	.21*	.27**	.27*
Transmissive teaching beliefs	-.07	-.06	.00	-.10
R^2	.08	.31**	.51**	.16*

$^{+}$ $p < .10$. * $p < .05$. ** $p < .01$.

University teachers' personal priority of teaching showed an effect on a multitude of observed teaching characteristics: It was linked to marginally better verbal and less visual structuring, more higher-order questions as well as higher observed student involvement and was also positively related to observed rapport. Notably, no significant effects could be seen in regard to teaching methods and the student-reported measures. The teaching beliefs showed the following result pattern: Whereas constructivist beliefs were not associated with any of the observed variables, they had positive effects on the student-reported scales concerning the clarity of teaching, the stimulation of student involvement, and rapport. Transmissive teaching beliefs were, on the other hand, not related to the student-reported measures, but showed a positive connection to visual structuring and a negative one to illustrativeness.

4.4 Discussion

This study advanced the research on university teachers and their impact on teaching in two ways: 1) It put forward the personal value lecturers attach to the teaching task as a central variable in the endeavour of explaining higher education teaching, and 2) it delivered detailed insight into the effects of teachers' teaching-related values and beliefs on their actual teaching practice, i.e. various specific teaching characteristics as observed by external raters and students.

4.4.1 Effects of the Task Value of Teaching and Teaching Beliefs

Overall, both the value attributed to the teaching task and the beliefs about university teaching showed relations with many of the investigated teaching characteristics, and thus, seem to be valuable predictors of higher education teaching. The effects were rather small in size. However, as the measurement of performance criteria by external raters is known to yield weaker relations (van Eerde & Thierry, 1996; Pritchard & Sanders, 1973; Schwab, Olian-Gottlieb, & Heneman, 1979; Turney, 1974), the obtained results can still be considered meaningful.

Reflecting on the results with regard to the four investigated aspects of teaching, it is first of all striking that none of the three predictors had a significant effect on the teaching methods used. So, it appears that neither diverging beliefs about teaching nor the individual teaching value is expressed in the choice of methods. In view of the literature striving to establish a link between the teachers' beliefs or approaches to the respective teaching methods used (e.g., Coffey & Gibbs, 2002), this finding is already quite noteworthy in itself.

Concerning the quality of instruction, the picture is more multifaceted: While the priority of teaching seems to be reflected in probably more verbal but less visual structuring, teachers with higher transmissive beliefs tend to pay special attention to the measures of visual structuring. At the same time, they seem to fall short with respect to the use of illustrative tools in their talks. Stronger constructivist beliefs come along with higher values on student-rated clarity. While the positive relations were in line with our hypotheses and can be explained by the teachers focusing on what they believe to be beneficial, the negative relations were unexpected and are not easily interpreted. Do teachers with transmissive beliefs, for instance, possibly regard examples and analogies as unnecessary?

Student involvement increased with both the priority of teaching and constructivist beliefs. The observed indicators were clearly associated with the priority of teaching, whereas the stimulation of involvement reported by the students was linked to constructivist beliefs. These positive effects seem plausible, as they may be achieved by better preparation and a focus on the students' learning.

A similar picture was to be seen with regard to rapport, the teacher's friendliness, concern, and respect for the students: Whereas the observed variable was positively related to the teachers' priorities, the student-rated variable pointed towards the beneficial effect of constructivist teaching beliefs.

In view of these findings, we conclude that the priority of teaching has an impact on diverse observable teaching characteristics, and proves to be an interesting teacher variable. Apparently, a high personal value of teaching is indeed mirrored in a university teacher's behaviour. Sometimes

this effect may actually be explained by the amount of effort put into the preparation of courses (cf. Cole et al., 2008; Dietrich et al., 2017; Turney, 1974). With criteria like rapport it seems more likely that it is the attitude itself which might be effective. Possibly the prioritizing of teaching corresponds to a higher degree of wholeheartedness in class and influences the way teachers relate with their students. Talking about university teachers' value of the teaching task, it is important to note that it is probably largely intrinsic in nature. University teachers commonly have a lot of freedom concerning the emphasis they put on their various work tasks, research often being considered more important and more fun than teaching (cf. Cretchley et al., 2014). There are no measures like promotion, raise in salary, or more job security based on the teaching performance. Thus, if lecturers indicate to regard teaching as their number one priority, it will most likely be connected to their self-esteem or their role picture (cf. Mitchell & Albright, 1972). This finding is in line with Turney's, who already discovered in 1974 that especially a high intrinsic task value enhances effort and performance.

Notably, the value of teaching and transmissive teaching beliefs exclusively showed relations to observed criterion variables, whereas constructivist beliefs were exclusively related to student-reported variables. This pattern was particularly remarkable regarding the aspects of student involvement and rapport. Here, the differently assessed indicators were highly correlated (see Table 4.2), indicating that the distinct measurements might indeed tap the same teaching aspect. We expected to find differential effects of the predictors with regard to the distinct aspects of teaching, but we did not anticipate the data source to be decisive. So, how to explain this result pattern? The two measurement modes, of course, differ in their respective qualities: The observers were trained to measure teaching characteristics as objectively as possible, and were – due to their job – in a position to compare many different university teachers even in distinct subject disciplines. Not having to focus on the learning content, they were further able to pay attention to single teaching characteristics. In combination with a sufficient interrater reliability, these arguments suggest a high validity of the observer ratings. The students, on the other hand, had a bigger sample of the specific teaching and thus a broader basis for their rating, as they attended more than just three sessions. During the course, they never payed attention to the various teaching aspects intentionally though, but were only asked to report them retrospectively at the end of the semester. Possibly, this last mentioned fact may have led to a more global evaluation of the teachers, resulting in an overall g-factor, which might reflect fundamental teacher traits. We hypothesize that university teachers with higher constructivist teaching beliefs may essentially stick out in that they focus more strongly on the students as the recipients of their teaching. This attitude might be "sensed" by the students and thus be the factor that ultimately influences their rating on diverse scales. Another explanation for the link between the student-reported variables and teachers' constructivist beliefs is related to the specific group of students present during the final survey. The possibility that courses of teachers with high constructivist beliefs systematically witnessed a stronger positive-selection of students towards the end of the semester, which may have entailed a more favourable evaluation of the teaching, cannot be ruled out. In any case, as the data source may have an impact on the findings, for future studies it seems advisable to include various data sources to check the reliability of the measured teaching variables.

4.4.2 Limitations of the Study

To be sure, as any empirical investigation, this study was impeded by several limitations: Of course, the effects are most probably not unidirectional. Based on theory, we presupposed primarily the effect of the teacher variables on their teaching. However, it has to be kept in mind that all the teachers in our sample had been teaching for at least three years. So their teaching experiences will have fed back into their personal convictions and beliefs, may have enforced certain priorities or weakened beginners' teaching enthusiasm. As our study has a correlational design, we refrain from drawing definite conclusions in terms of causality.

The sample size of 79 university teachers certainly restrained the analytical options of this study. For example, it denied the inclusion of further control variables such as the course level, so that potential further influences on higher education teaching could not be considered. Similarly, it did not allow for a more specific consideration of the subject discipline. A greater differentiation might have permitted more informative results pertaining to the role of distinct subject cultures.

Further potential weaknesses of the study concern the measurement of some variables: We assessed our principal predictor of value only dichotomously as priority of teaching. As our results show, this rather rough measurement still yields interesting findings. Yet, we cannot rule out the possibility that a more fine-grained assessment of the value attributed to the teaching task might have led to more and possibly stronger effects. Apart from that, two of the observed teaching characteristics – visual structuring and rapport – were assessed with a rather low interrater reliability. We decided to include them nevertheless, as they are important indicators for two central aspects of teaching. Plus, they also seem to be measured well enough to still show relations with the other variables.

Lastly, we want to add that this study was largely exploratory in nature. Our results may flash some light into the black box of teaching in higher education, but they need to be affirmed by future studies.

4.4.3 Conclusion

So, what's the bottom line? 1) It may be more decisive whether university teachers place a high priority on the task of teaching rather than what exactly they think of how it should be done. Our results suggest that various aspects of teaching – the quality of instruction, the student involvement, and rapport – are improved if a university teacher prioritizes teaching. At the same time, teaching beliefs may not be unimportant. Especially constructivist beliefs had positive effects on the students' perception of the teaching. In view of a possible survivor bias, these effects have to be interpreted with caution, though.

2) Seeing the source specificity in the effect pattern, our results present a strong argument for multiperspective research in higher education. Especially studies investigating the effects of teacher characteristics that may contribute to a higher student acceptance (e.g., enthusiasm, cf. Fischer & Hänze, 2019) on teaching behaviour and other outcome measures are well-advised to not only rely on student-reported data. Observational measurement, for instance, provides a good opportunity to avoid distortions in teaching evaluations that come with student reports.

3) Last, but not least: How are the findings relevant for practice and real life? To advance higher education teaching, next to attempts to favourably alter university teachers' beliefs, it may prove beneficial to also consider task value. Means to increase the value of teaching for university teachers could of course be extrinsic in nature, e.g., monetary incentives or public recognition of good teaching. However, pursuing this approach, the problem of how to economically and objectively assess the quality of higher education teaching arises – not even to mention the normative decision of what may be considered high quality teaching. The augmentation of the intrinsic value of the teaching task might not only be more effective but also better feasible – by shaping the university teachers' perception of their role as instructors, by emphasizing their relevance for society in educating the coming generation of professionals, and by a pronounced appreciation of their teaching effort. No matter how this might be achieved, a high priority of teaching on the teachers' side may result in better tertiary education.

5 Substudy on the Process Dimension of Higher Education Teaching [7]

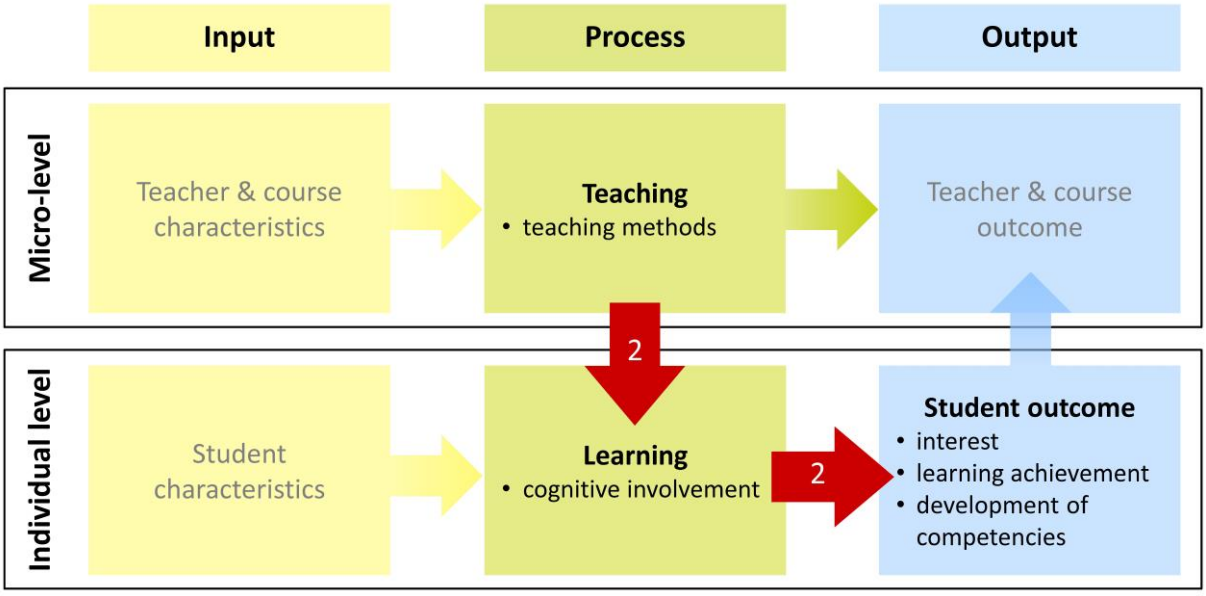

Figure 5.0. Illustration of the second substudy within the framework of higher education teaching.

[7] This chapter was co-authored by Martin Hänze and published in the International Journal of Educational Research. Please cite:
Fischer, E., & Hänze, M. (2019). Back from "guide on the side" to "sage on the stage"? Effects of teacher-guided and student-activating teaching methods on student learning in higher education. *International Journal of Educational Research, 95*, 26-35. doi.org/10.1016/j.ijer.2019.03.001 (open access)
As an exception, American English was used in this paper. Minor modifications (e.g., numbering of tables and headings) were made here to increase consistency within the dissertation. Sources are cited in a way that also allows for separate reading (citation of all authors with first mention in chapter).

Back from "Guide on the Side" to "Sage on the Stage"?
Effects of Teacher-Guided and Student-Activating Teaching Methods on Student Learning in Higher Education

Abstract

This field study compares the effectiveness of teacher-guided and student-activating teaching methods. Expert observations of 80 university courses were combined with self-report data from 1,713 students attending the courses. Controlling for students' initial interest on the individual level and for course format, homework, and initial interest on the course level, two-level path analyses with the amount of teacher-guided and student-activating methods as predictors, and students' final interest, subjective learning achievement, and perceived development of academic competencies as criteria – all mediated by the students' cognitive involvement – revealed opposing effects of the two methods. Teacher-guided methods were associated with an increase in students' cognitive involvement, interest, learning achievement, and development of academic competencies, whereas student-activating methods tended to show negative effects.

Keywords: constructivist learning theories; higher education; student-activating teaching methods; teacher-guided teaching methods

5.1 Introduction

In many European countries, student-activating methods in higher education teaching are currently being promoted strongly. The German Council of Science and Humanities (2008, p. 62), for instance, has called for university students to be teachers' active partners in the learning process rather than passive receivers of learning material. These recommendations are reinforced by considerable financial incentives. For the period 2011-2020, the Federal Ministry of Education and Research invested two billion Euros toward improving student support and teaching quality in higher education. This exemplary national policy is interlinked with efforts in the field of higher education at the European level. Fostering student-centered learning became an explicit aim of the Bologna process during the follow-up conference in Leuven in 2009 (EHEA Ministerial Conference, 2009, p. 3 f.). In the Bucharest Communiqué (EHEA Ministerial Conference, 2012) three years later, the ministers again emphasized their commitment to "promote student-centred learning in higher education, characterised by innovative methods of teaching that involve students as active participants in their own learning" (p. 2). According to the European Students' Union (2015), one of the principles of student-centered learning is a preference for enabling over telling: "In simply imparting facts and knowledge to students (telling) the initiative, preparation and content comes mainly from the teacher. The SCL [student-centred learning] approach aims to give the student greater responsibility enabling the student to think, process, analyse, synthesise, criticise, apply, solve problems, etc." (p. 7). In particular, the goal of employability and the corresponding focus on acquisition of skills within the Bologna process led to a call for new teaching formats (e.g., German Rectors' Conference, 2013; Schaper, 2012). According to an expert report on competence orientation in higher education (Schaper, 2012) prepared for the German government, competencies are not gained through receptive learning, but require active, hands-on, and problem-oriented engagement with the learning material. Activating teaching formats and situated tasks are not only necessary prerequisites for developing competencies but also entail higher cognitive involvement and deeper processing of the learning contents (p. 56).

The described policy development is in line with the upward trend of the constructivist learning approach. In this approach, student-activating learning formats are believed to lead to deeper understanding, stronger motivation, and greater development of competencies, whereas traditional teaching formats are thought to produce inert knowledge and thus to be less effective. Under the catchy title "From Sage on the Stage to Guide on the Side" King (1993) called for more active, mostly cooperative learning formats in college classrooms as early as 1993. Not every scholar is, however, convinced that these strongly supported and publicly advocated learning formats really render the ascribed benefits. For instance, Tobias (2009) observed about the related literature: "There is stimulating rhetoric for the constructivist position, but relatively little research supporting it" (p. 346).

This study sought to contribute to the empirical evidence on the effects that student-activating and teacher-guided methods have on student learning in everyday university teaching.

5.1.1 Theoretical Debate on Constructivism

There is a long-standing debate on the concept of constructivism and its implications for pedagogical practice (e.g., Renkl, 2009; Tobias & Duffy, 2009). The debate originates in diverging understandings and thus inconsistent use of the term *constructivism* and is spurred on by two almost opposing views. One group of scholars thinks of constructivism more as an approach for student-centered teaching and is concerned mainly with specific learning formats (cf. Loyens & Gijbels, 2008; Steffe & Gale, 1995), whereas another faction of scholars puts forward that constructivism is mainly a theory explaining the cognitive mechanics of learning and does not yield immediate conclusions for recommendable teaching methods (cf. Mayer, 2009; Renkl, 2009).

Constructivism as an Approach for Student-centered Teaching

The foundations of constructivism are located in the educational theories of Vygotsky (1987) and Piaget (1952), among others. As the more immediate stimulus for the ascent of constructivism in pedagogy, Tobias and Duffy (2009) point to articles emphasizing that learning is situated and knowledge is a product of an activity in a specific context (e.g., Brown, Collins, & Duiguid, 1989) or promoting the idea that "natural" learning involves socially shared activities, direct engagement, use of cognitive tools, and development of specific skills needed in that particular situation (e.g., Resnick, 1987). The central claim of these constructivist approaches to learning is that learners build up their knowledge themselves by individually discovering and transforming complex information, checking new information against old rules, and revising prior schemata if necessary. Great emphasis is put on students being active learners who are best supported by different forms of student-centered instruction (e.g., discovery learning, cognitive apprenticeships) that work with authentic learning tasks and complex problems rather than simplified ones (e.g., Duffy, Lowyck, & Jonassen, 1993; Loyens & Gijbels, 2008; Steffe & Gale, 1995). The presumed benefits of these learning approaches include greater sustainability and higher transferability of knowledge and competencies as well as motivational advantages.

According to this understanding of constructivism, student-centered, activating learning environments are more likely to initiate student cognitive involvement than are traditional, teacher-centered settings. High-level cognitive processes are a precondition for desirable learning outcomes such as interest or learning achievement. The acquisition of competencies is deemed particularly dependent on actual experiences, hands-on tasks, project work, or problem-based learning. Traditional learning formats are thought to lead to passive learning, resulting in inert knowledge at best.

Constructivism as a Theory of Learning

In contrast, Mayer (2009) makes a case for constructivism as a theory of learning rather than as a prescription for instruction. In his understanding, the theoretical core claim of constructivism is that the learner builds mental representations by cognitively processing new information during learning. For Mayer, the "idea that constructivist learning is caused by active methods of instruction rather than by active learning" is the "constructivist teaching fallacy" (p. 188). He stresses that high cognitive activity during learning is not dependent on high behavioral activity. Thus, it may be inappropriate to assume that active cognitive learning requires teaching methods that promote

hands-on behavioral activity; it may also be inappropriate to assume that passive instructional methods fail to promote active cognitive learning: “Behavioral activity during learning does not guarantee that the learner will engage in appropriate cognitive processing, and behavioral inactivity during learning does not guarantee that the learner will not engage in appropriate cognitive processing” (p. 185). As part of the ongoing debate, Renkl (2009) pointedly reminded the research community of what, as he sees it, constructivism actually claims – that all learning happens by means of knowledge construction on the part of the learner. There is no other way but for the learner himself to take in the new information, often presented in written or spoken texts, interpret it on the basis of prior knowledge, and integrate it into existing knowledge networks. Thus, every learning environment, if effective at all, will inevitably be constructive; there is no such thing as a non-constructive learning environment.

A growing body of literature aligning with this understanding of constructivism tries to deduce instructional designs from human cognitive architecture. Building on the theory of cognitive load, for instance, instructional techniques like spaced practice, retrieval practice, or worked examples are proposed (cf. Sweller, Ayres, & Kalyuga, 2011). In higher education, learning could be supported by a well-structured presentation of the learning material and by pointing out connections to prior knowledge. This might make it easier for learners to integrate the new information into their existing schemes, which may lead to more cognitive links, a better network, and deeper understanding. However, regarding global approaches like student-activating or teacher-guided settings, scholars refrain from deciding which of the two more strongly supports the construction of new knowledge (Mayer, 2009; Renkl, 2009). It is conceivable that a thoroughly prepared lecture is effective, but also in learning formats where the students themselves become active individually or in a group, they might co-construct knowledge together, or apply and thereby deepen their knowledge. None of the learning formats can guarantee that certain cognitive processes will occur, though.

Thus, there is a consensus between the two branches of constructivism that knowledge is built up actively by the learner; but there is disagreement concerning conclusions for pedagogical practice and the teaching methods to be used.

5.1.2 Empirical Evidence for and against the Efficacy of Student-activating Methods

The question of what instruction formats are most effective in higher education triggered empirical research very early on (McKeachie, 1990). Krumboltz and Farquhar (1957) investigated instructor-centered, student-centered, and eclectic teaching and found no significant differences in student learning achievement but diverging results for motivation: Students in the eclectic courses were most highly motivated, the students in the instructor-centered courses ranged second, and the students in the student-centered courses showed the least increase in motivation. A decade later, Webb and Baird (1968) compared traditional teacher-centered instruction to student-centered instruction. Students taught in the student-centered course design had significantly better test scores than the control group with teacher-centered instruction.

Today, a vast number of studies support the superiority of student-centered, activating learning formats in tertiary education. Deslauriers, Schelew, and Wiemann (2011), for example,

examined the benefits of teaching approaches involving challenging questions, reasoning, and problem solving in a large-enrollment introductory physics class. Student-activating methods led to higher engagement and much better learning than traditional lecturing. Freeman et al. (2014) conducted a meta-analysis with studies in the STEM disciplines and presented strong evidence for the predominance of activating learning settings in comparison to traditional lecturing: Grade scores were higher and failure rates considerably lower in active learning courses than in traditional ones. These results were robust over various disciplines and course sizes. Dochy, Segers, Van den Bossche, and Gijbels (2003) and Schmidt, van der Molen, te Winkel, and Wijnen (2009) conducted meta-analyses on studies investigating problem-based learning (PBL), a specific form of student-activating teaching that involves small group learning under the guidance of a tutor, work on authentic problems, and self-directed learning. Dochy and colleagues found robust positive effects of PBL on student skills, but a tendency toward negative results for student knowledge. Schmidt and colleagues focused on PBL in medical education and presented similar findings: Students performed much better in professional skills, i.e., interpersonal and practical medical skills, and drop-out rates as well as study time were lower in PBL programs. There were, however, only small and inconsistent differences with regard to medical knowledge and diagnostic reasoning. Tynjälä (1999) conducted a complex qualitative study to compare learning differences in students who studied the course material in a "constructivist" learning environment with regular writing assignments and group discussions and students who were taught in a traditional lecture setting. Although there were no differences between the groups in knowledge acquisition, students in the constructivist learning environment seemed to have acquired more diversified knowledge. When asked about their learning experience, they mentioned more frequently that they had developed their thinking and acquired skills during the course. An interesting approach to investigate the effect of activating teaching methods was taken by Cherney (2008). Instead of comparing groups of students in differing learning environments, Cherney compared undergraduate students' free recall for course content delivered in differing ways and found that the items that were listed most frequently were concepts introduced through active learning exercises.

In contrast to these findings, however, a number of studies either found advantages of traditional teaching approaches or did not yield any differences at all between the instruction modes. In the attempt to show that student-centered learning environments encourage a deep approach to learning, for example, Baeten, Struyven, and Dochy (2013) compared a lecture-based and a student-centered learning environment. Lecture-based instruction did not lead to a change in the use of the deep approach to learning and brought about only a small increase in the use of surface technics. The student-centered learning environment led to less use of deep approach strategies and pushed students towards a surface approach. Comparing lecture-based and student-activating courses with the same content, Struyven, Dochy, and Janssens (2008) investigated the impact of students' tastes in teaching methods on perceived quality of the learning environment, their own learning, and their actual performance. The study found that lecture-taught students were all quite content with their course, whereas the opinions on student-centered teaching diverged quite strongly – students felt either extremely positive or extremely negative about it. The students' likes and dislikes regarding instruction reaped consequences and had a positive effect on their evaluation of the learning

environments, their own learning, and their performance. As students liked lectures best, lectures were also attributed the highest quality, best teaching, clearest goals, and most appropriate workload; even on the scales on learning generic skills and independence in learning, lectures scored as high as the highest student-activating course. Furthermore, student performance was also best in lectures; only the student-centered course with a multiple-choice test led to similar results. In another study, Loveland (2014) for two semesters divided an introductory university-level statistics course into two sections – one section was taught entirely in traditional lecture style, the other using active learning methods and a minimal amount of lecturing. Analyzing student exam scores and their attitudes, Loveland found no significant differences in outcomes: The activity-based teaching method did not lead to higher student comprehension or procedural ability, nor did it lead to more positive student attitudes. The student comments indicated a positive response to the activity-based methods but also a desire for more teacher-centered time in the activity course.

Thus, no clear picture emerges from a look at publications on student- and teacher-centered instruction in higher education. This is partly due to the multitude of learning outcomes investigated. For example, student-activating teaching seems to really support the development of practical skills – at least in the PBL-context, but its effect on motivational and attitudinal student criteria is indistinct. With respect to measures of knowledge acquisition in particular, the findings are inconsistent and cannot substantiate the assumption that activating learning entails better understanding.

A challenge connected with the research on teaching methods is the fact that teaching effectiveness will always depend on the concrete implementation of any instructional approach. It is the appropriateness and the quality of realization of a method that is ultimately decisive. So, when comparing different methods, there is always the question of whether the implementation quality really was the same. As a number of studies set out to demonstrate the advantage of one teaching approach – currently mostly student-centered methods, it may sometimes be questionable whether the teaching format in the control condition was implemented as thoroughly as the one under investigation.

5.1.3 Implications for this Study and Research Question

While there are numerous studies examining the influence of different teaching approaches on various outcome variables in higher education, little research is available on the immediate effect of the teaching methods – the process of learning, the construction of new knowledge itself. Several studies investigated the mediating effect of student engagement in higher education (e.g., Gasiewski, Eagan, Garcia, Hurtado, & Chang, 2012; Zumbrunn, McKim, Buhs, & Hawley, 2014). However, as engagement is mostly conceptualized as a broad construct covering classroom participation as well as the use of support systems like the professors' office hours or supplemental courses (cf. Sinatra, Heddy, & Lombardi, 2015; Fredricks, Filsecker, & Lawson, 2016), these studies hardly offer conclusions about the actual learning process. What is more, visible participation is more likely to occur in small courses, as they provide more opportunity for students to become involved behaviorally (e.g., by doing an exercise). One facet of engagement can equally occur in lectures and may thus be less confounded with the course format: student cognitive involvement. It is also applicable in various disciplines, whereas other measures of engagement may strongly depend on

the specific “culture” or “logistics” of a certain subject. To advance the discussion on effective instruction modes based on constructivism, empirical evidence is needed that can trace student success in the outcome variables to their cognitive involvement in class, which was in turn triggered by the respective teaching methods.

In contrast to the majority of studies investigating the effectiveness of teaching approaches, the study presented here does not evaluate student-centered courses that were specifically designed for the purpose of comparison to traditional lectures. It is a field study analyzing “real life” university teaching as it commonly occurs. The subjects of comparison are not entire courses but the amount of certain methods used within each of the university courses investigated. The sample thus contains big lecture courses with activating elements and smaller student-centered seminar courses with teacher-guided parts.

To evaluate learning achievement, we employed two distinct measures – one to assess domain-specific knowledge gains and the other to detect the development of academic competencies, i.e., general abilities such as reasoning, critical thinking, or problem solving, all implicit goals of most university courses. According to Shavelson (2010, p. 18), both of these measures of learning outcome should be considered in assessing learning. Furthermore, our study includes interest as a motivational criterion, since activating methods are often said to especially boost student motivation.

Following the basic constructivist claim that the students’ cognitive involvement is the decisive factor for any learning outcome, this paper suggests a mediation model with student cognitive involvement as a precondition for all outcome variables.

Our research question was: How do teacher-guided and student-activating methods in higher education affect student interest, learning achievement, and the development of academic competencies in a field setting, and does student cognitive involvement mediate the effects?[8]

5.2 Method

5.2.1 Sample

The sample was obtained at a middle-sized, public university in Germany and consists of 80 courses and the students enrolled in them. The courses were 48 lecture and 32 seminar courses in various academic disciplines, ranging from physics to social sciences, languages, and music: Twenty courses fell into the category of the natural sciences and mathematics, and 60 were in other disciplines. The courses usually consisted of 14 sessions which usually had a duration of 90 min each. The course teachers were asked personally to take part in the study with the incentive of receiving feedback on their teaching; the participation rate was 44%. Of all students enrolled in the courses,

[8] We prefer not to use the terms “teacher-centered” and “student-centered teaching” in our study, as the teaching provided by an instructor – in any modus – is usually meant to be student-centered. To label our constructs we prefer a terminology that describes manifest, observable occurrences rather than terms that prescribe or interpret them. In this paper, therefore, the term “teacher-guided methods” refers to methods where the instructor is in the foreground and actively steers students’ learning. “Student-activating methods” comprise all learning formats that activate all students at the same time (e.g., an exercise, a reading task, group work).

1,716 students took part in the entry survey as well as the final survey; 1,713 of them have values on the exogenous student variable and are thus included in the analyses for this study. These students (59% female) were on average *M* = 23.2 years old (*SD* = 4.3) and had been studying for *M* = 4.2 (*SD* = 2.6) semesters. They made up 35.2% of all students who took part in the entry questionnaire and differed only slightly from students who did not participate in the final survey. The differences in the personal characteristics and preconditions of the students included and excluded in the analyses were small (max. Cohen's *d* = .21, conscientiousness). Participation was voluntary for both the teachers and students.

5.2.2 Procedure

The study design was correlational and longitudinal. To assess student preconditions, the students filled out an entry survey in each course during the first two weeks of the semester. At the beginning, in the middle, and at the end of the semester, trained observers visited each course to identify the teaching methods used. The final survey was conducted in the last two weeks of the semester; students were asked to describe retrospectively the course and their own engagement and to evaluate their learning outcomes. The data was gathered during three semesters in the years 2014-2015.

5.2.3 Instruments

Predictors (course level)

To assess the teaching methods employed in the various courses, trained observers inspected the courses (unannounced) three times, using a standardized rating form to describe each session. The observers were student research assistants who had completed approximately 24 hours of training. About 25% of the visits were done jointly to determine the degree of agreement between the raters. Interrater reliability was calculated separately for each semester; the intraclass correlation coefficients ($ICC_{1,1}$) were satisfactory (see below). The three rating scores were averaged for each course. The observers documented the exact amount of time that distinct teaching methods were employed in the classroom. The measure used for the predictors in this study was the share of the total teaching time that was used for the particular type of teaching method. All observed teaching methods were grouped into different categories: The category *teacher-guided methods* included talks by the teacher or a guest speaker, the use of music or videos, and demonstrations by the teacher such as calculations or experiments. The reliability of the three measurements assessed by Cronbach's alpha was .93, indicating a high stability; the $ICC_{1,1}$ for this category in the three semesters were .88, .98, and .81. The category *student-activating methods* included methods that intend to activate all students (as opposed to only a few) without direct involvement of the teacher. Examples are reading tasks, exercises, pair- and group-work, or games. The reliability of the three measurements assessed by Cronbach's alpha was .73, indicating a high stability as well; the $ICC_{1,1}$ for this type of teaching methods were .58, .97, and .93. These two contrasting modes of instruction excluded methods of teacher-student interaction (such as discussions or text work with the teacher) as well as student-guided methods (such as student moderation, presentations, or micro-teaching), where only few students are directly activated.

Criteria and mediator (student level)

We assessed the three criteria – students' final interest, subjective learning achievement, and perceived development of academic competencies – and the mediator – cognitive involvement – at the end of the semester in the final survey (paper and pencil or online version). We developed a six-item scale for *cognitive involvement* that encompassed cognitive aspects of behavioral engagement (cf. Jang, Kim, & Reeve, 2016), focusing on student attentiveness in the sessions and elaborative processes in the aftermath (α = .82, e.g., "During this course I was almost always thinking along"; "Sometimes I reflected on the learning content even after the course"). *Final interest* was measured with a five-item scale (α = .86, e.g., "I found many of the topics covered in this course very interesting") aligned with the central aspects of interest theory: high subjective relevance, association with positive feelings, and intrinsic character (cf. Schiefele, 1991). *Subjective learning achievement* was evaluated with five items asking about the gain in domain-specific knowledge (α = .83, e.g., "I learned a lot in this course"). The third criterion, perceived *development of academic competencies*, was measured by a 14-item scale; the items were derived from statements made by the participating university teachers in preliminary interviews when asked about their teaching goals. The construct had five facets: problem solving, reasoning and arguing, interrelating, adopting multiple perspectives, and elaborating (α = .93, e.g., "I learned here to examine something from various perspectives or using distinct theories"; "I was encouraged to scrutinize data, assumptions, or the like critically"). The wording of the scale was similar to the wording of the generic skills scale of the Course Experience Questionnaire (Wilson, Lizzio, & Ramsden, 1997) and the knowledge processing scale of the instrument by Braun and Leidner (2009); the content also resembled the cognitive skills scale of the Personal and Educational Development Inventory (Lawless & Richardson, 2004). As recommended by Braun, Woodley, Richardson, and Leidner (2012), all four scales referred explicitly to the courses under investigation. Answers were given on a 6-point Likert-scale (1 *disagree*, 6 *fully agree*).

Control variables (student and course level)

To account for the influence of the students' preconditions – both on the course and on the student level, their interest in the course was assessed in the entry survey and used as a control variable: *Initial interest* (Schiefele, 1991) was measured with a seven-item scale (α = .89, e.g. "I find many of the topics covered in this course very interesting"); answers were again given on a 6-point Likert-scale (1 *disagree*, 6 *fully agree*). To be able to investigate simultaneously teaching methods in distinct types of courses without neglecting their organizational differences, we also included the control variable *course format* (dichotomous: 0 *lecture*, 1 *seminar*). In Germany, lectures often imply large student numbers and teachers imparting a broad subject matter, whereas seminars usually come with small learning groups and an elaboration of more specific learning content with stronger student participation. However, courses of both formats vary in size and learning arrangement, blurring the lines. Lastly, a supplementary didactic feature of university courses that is meant to influence the students' cognitive involvement was included in the analyses to bring about the specific effects of the teaching methods: *homework*, i.e., obligatory learning activities outside of class (dichotomous: 0 *no*, 1 *yes*).

5.3 Results

Teacher-guided methods were found in almost every course (n = 77) and made up M = 64.0% (SD = 34.5) of the total teaching time in that subsample. Student-activating methods were less prevalent (n = 32) and, if employed at all, consumed on average M = 13.8% (SD = 9.2) of the teaching time.

Over all the courses, the students indicated average cognitive involvement of M = 4.1 (SD = 1.0). The average final interest amounted to M = 4.1 (SD = 1.1), mean subjective learning achievement was M = 4.5 (SD = 1.0), and development of academic competencies was indicated as M = 3.2 (SD = 1.1) on average.

To test our theoretical assumptions, we ran a two-level path model with the two teaching methods as predictors, and final interest, subjective learning achievement, and perceived development of academic competencies as criteria, all mediated by student cognitive involvement, using Mplus 7 (Muthén & Muthén, 1998-2014). Direct paths to the criteria were omitted; the model fit values assent to this decision. The $ICC_{1,1}$ of the dependent student variables, which indicate the agreement of the students within their course, ranged between .16 and .24, confirming the appropriateness of the multi-level approach. Figure 5.1 displays the results.

Teacher-guided methods had a positive effect on student cognitive involvement. Student-activating methods, however, showed with p = .07 a marginal negative effect. Cognitive involvement in turn had strong positive relations with all three outcome variables, both on the individual and the course level, and thus served as a mediator.

As the main model featured rather high correlations between several predictor variables and as multicollinearity may lead to unreliable results, successive analyses were run to separately investigate the two teaching methods in lectures and seminars respectively (cf. Figures 5.2 and 5.3). To support the analyses in the smaller seminar sample, two paths, which were not significant in the main model, were fixed to zero.

The effect of teacher-guided methods remained stable in lecture courses but did not reach significance in seminars. The negative effect of student-activating methods, in contrast, was only marginal in lectures but significant in seminars.

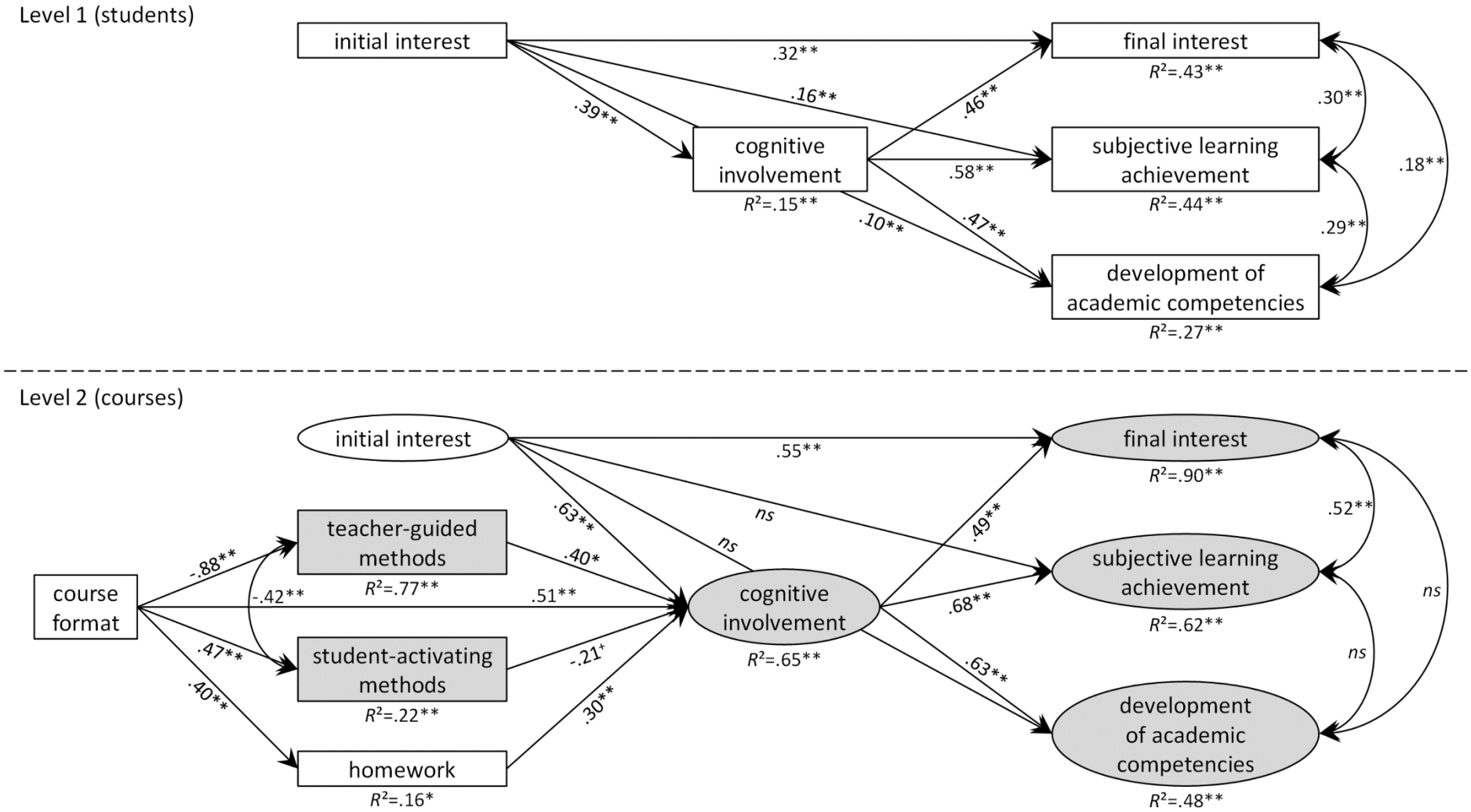

Figure 5.1. Two-level path model of effects of teacher-guided and student-activating teaching methods in higher education on student learning outcomes mediated by student cognitive involvement with standardized coefficients; highlighted in grey is the theoretical model under investigation. $\chi^2(17) = 38.5$, $p = .00$; RMSEA = .03; CFI = .99; TLI = .98.
$^+p < .10$. $*p < .05$. $**p < .01$.

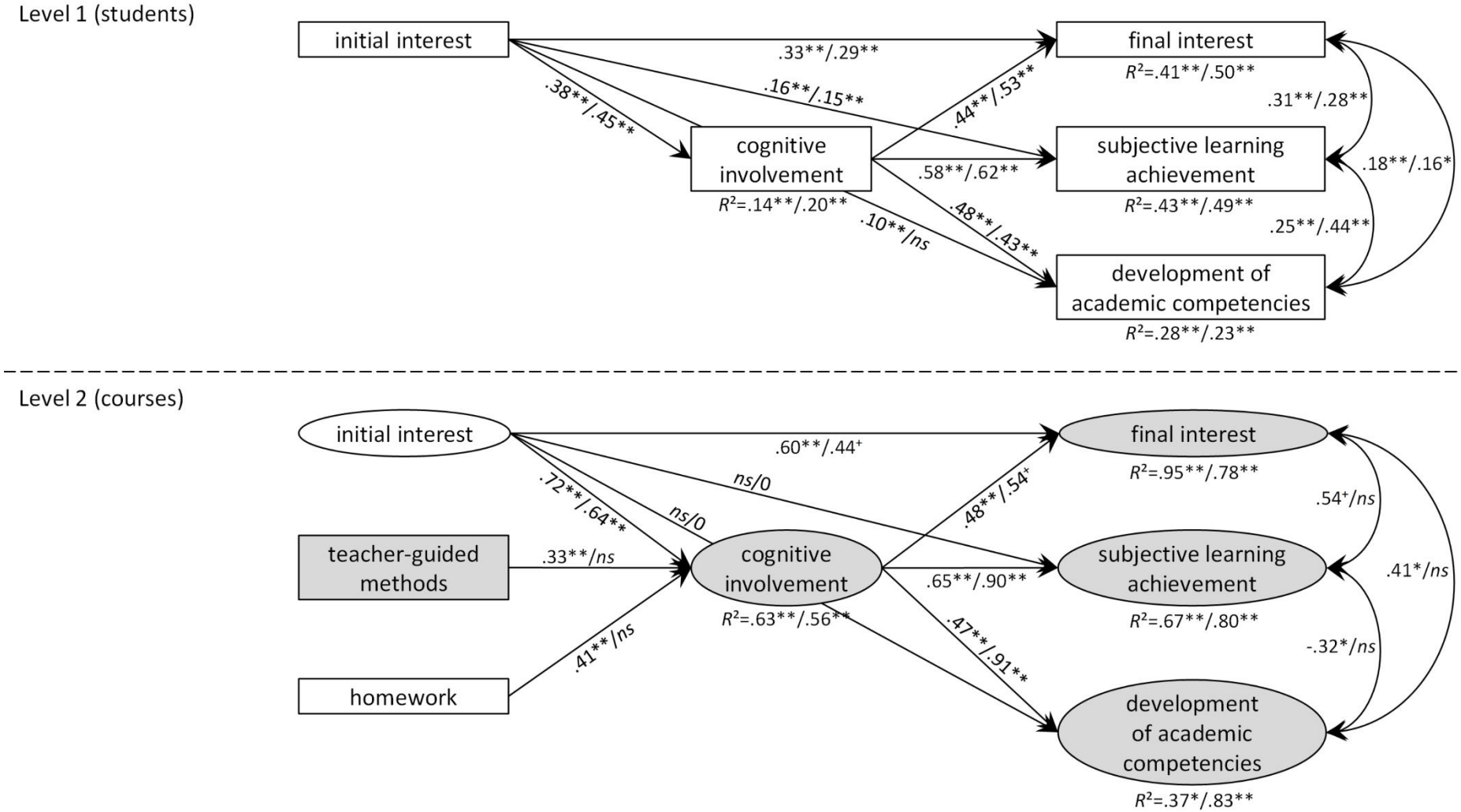

Figure 5.2. Two-level path models of effects of teacher-guided methods in higher education on student learning outcomes mediated by student cognitive involvement with standardized coefficients, computed for each course format separately (lectures/seminars); highlighted in grey is the theoretical model under investigation. Model fit for lectures: $\chi^2(6) = 15.9$, $p = .01$; RMSEA = .04; CFI = 1.00; TLI = .98. Model fit for seminars: $\chi^2(8) = 10.6$, $p = .22$; RMSEA = .03; CFI = 1.00; TLI = .99.

$^{+}p < .10$. $^{*}p < .05$. $^{**}p < .01$.

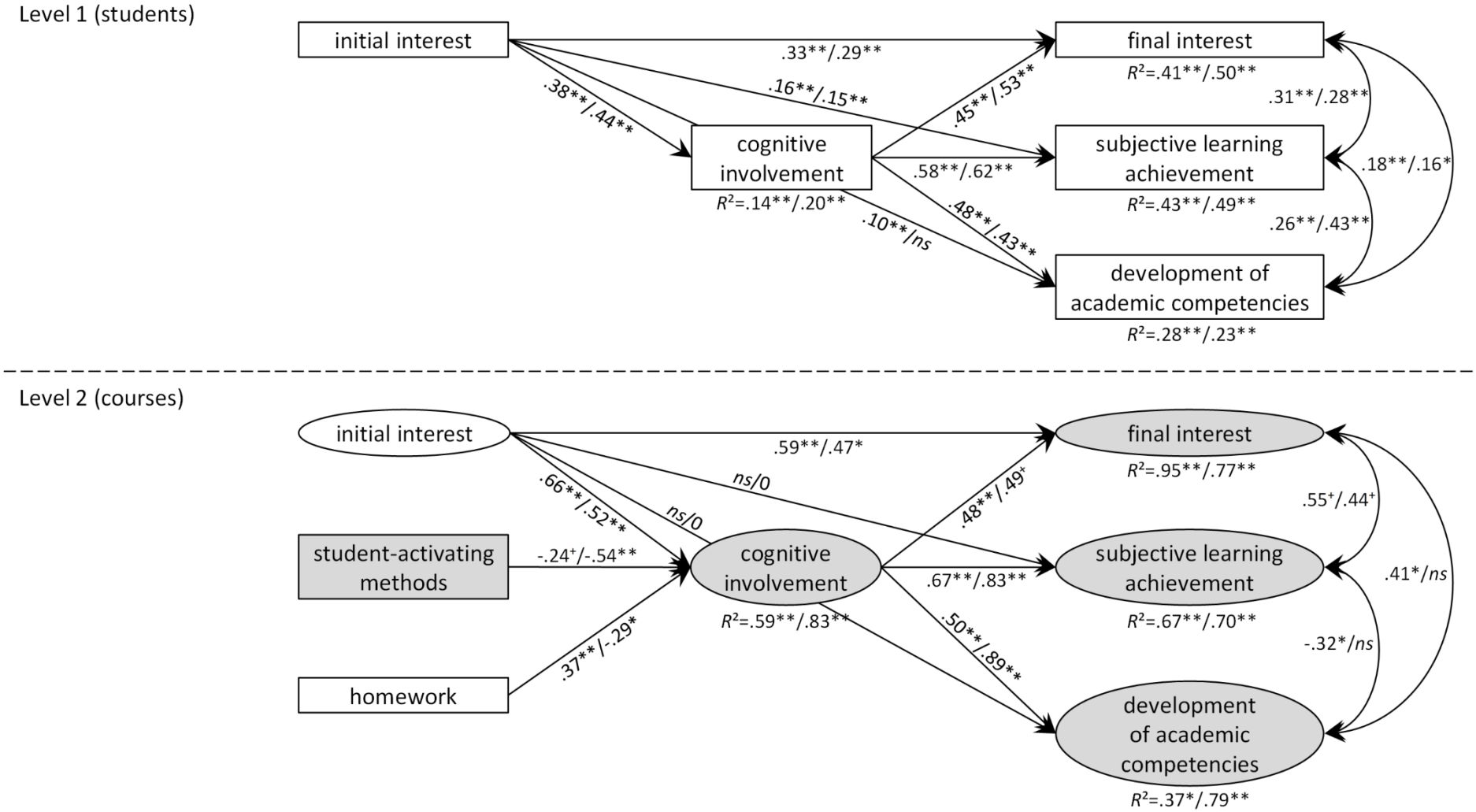

Figure 5.3. Two-level path models of effects of student-activating methods in higher education on student learning outcomes mediated by student cognitive involvement with standardized coefficients, computed for each course format separately (lectures/seminars); highlighted in grey is the theoretical model under investigation. Model fit for lectures: $\chi^2(6) = 10.7$, $p = .10$; RMSEA = .02; CFI = 1.00; TLI = .99. Model fit for seminars: $\chi^2(8) = 7.6$, $p = .47$; RMSEA = .00; CFI = 1.00; TLI = 1.00.

$^+p < .10$. $*p < .05$. $**p < .01$.

5.4 Discussion

This study investigated the effect of teacher-guided and student-activating teaching methods on student cognitive involvement, interest, learning achievement, and academic competencies in higher education. In accordance with the basic idea of constructivism – that learners build up new knowledge themselves – a mediation model was proposed with cognitive involvement conveying the effects of the teaching formats to the outcome variables. To increase the generalizability of the results, the study included university courses, both seminars and lecture courses, across disciplines. To further ensure ecological validity, the study refrained from comparing courses specifically designated as innovative to deliberately traditional courses in an experimental setting and instead investigated real life teaching at a middle-sized, public university. Expert ratings of the teaching methods employed were combined with student self-reports at the beginning and at the end of the semester, and two-level path analyses were carried out.

The results indicate oppositional effects of the amount of teacher-guided and student-activating methods in higher education teaching on student learning. While the use of teacher-guided methods seems to promote cognitive involvement – at least in lecture courses, the use of student-activating methods tends to have negative effects. This result complements the findings of previous studies (e.g., Baeten et al., 2013) but contradicts the research supporting the superiority of student-activating methods.

The suggested mediation model can be approved: The data conform to the hypothesis of cognitive involvement transmitting the effects of teaching methods to the outcome variables. Hence, the results are in line with constructivist learning theory regarding cognitive involvement as the prerequisite for learning. Only if a learner is actually thinking along and elaborating new information may he or she adopt higher interest, acquire greater learning achievement, and develop academic competencies.

5.4.1 Effectiveness of Teacher-guided and Student-activating Methods

With regard to the positive association between teacher-guided methods and student cognitive involvement, it must first be stated that it is indeed misleading to label learning as *passive* in these learning formats. At the same time, it can be questioned whether behavioral activity indicated by the amount of time spent on student-activating methods actually supports or instead hampers cognitive involvement. Apparently, high behavioral activity does not automatically lead to cognitive activity (cf. Mayer, 2009). Concerning the debate on teaching approaches in higher education and other settings, we therefore advocate careful and clear use of language. From our point of view, the distinction between passive and active learning is not helpful. We agree with Renkl (2009) that learning is always an activity per se; there is no such thing as passive learning. And regarding the categorization of instructional settings, we also recommend avoiding the use of *passive* as an attribute. Listening is also an action, and it can prove conducive to learning. Whether a learning format involves hands-on activities or more cognitive reconstruction on the part of students should be expressed by describing what is actually happening without inferring the respective learning process (e.g., student-activating methods, teacher-guided methods).

Now, how can the effects of the two distinct teaching methods on the criteria be explained? The growth of student interest in courses with longer teacher-guided phases might be ascribed to the greater appearance of experts who may raise interest by somehow transmitting their own enthusiasm for the subject matter (cf. Patrick, Hisley, & Kempler, 2000). A possible explanation for the positive influence of the teacher-guided methods on the perceived development of academic competencies is Banduras' (1977) theory of model learning. Here, the learner re-enacts or reproduces observations of a role model, which may lead to a change of attitude or behavior. In higher education, students could thus learn and adapt to the academic way of thinking and arguing by listening to and observing their teachers. Whereas hands-on approaches in teaching might indeed be required for the development of practical abilities, the scientific thinking and reasoning might also be acquired – and possibly more effectively – in lecture settings through model learning. Learning how to think and reason academically by observing a "sage on the stage" during teacher-guided methods seems more likely than by working alone or with fellow students during activating phases, where the teacher operates only as a "guide on the side".

However, the results counter to the activating methods and in favor of lecture-style teaching may also partly be explained by students' personal likes and dislikes regarding teaching formats (cf. Struyven et al., 2008, see also Hativa & Birenbaum, 2000). It is well conceivable that a certain sense of appropriateness about how to study at university is decisive for acceptance and evaluation of any learning format. Activating teaching methods might be rejected just because they are not well liked or deemed inappropriate for higher education, in the sense of, "We are not school children anymore!" But even if the diverging effects of the distinct teaching approaches were indeed primarily caused by the students' taste and not the method itself, this dynamic would need to be considered when discussing the improvement of higher education teaching. The best learning format is worth nothing if students are unwilling to use it.

Another explanation for the findings lies in the question of teaching quality. Higher education teachers might be the most comfortable with teacher-guided methods, since giving talks is part of what they learn to excel at when pursuing a scientific career. Specific teacher training that might provide guidance on using activating methods in teaching is still not widespread in many countries. Hence, the quality of the well-known methods might be better. Following this path of argumentation, however, it must be kept in mind that if alternative teaching formats are being used at all, it is likely by people who are convinced of their usefulness and who try to implement them the best they can.

However, in regard to the above interpretations of the study's findings, it should be remembered that all the effects on the course level only refer to the part of variance that is explained by the level two variables (between 16% and 24%). Thus, when explicating the opposing effects of teacher-guided and student-activating methods, practical significance should not be overrated.

5.4.2 Limitations of the Study

A number of limitations restrict the explanatory power of the study. First of all, the correlational design does not allow for straightforward causal interpretation. We attempted to rule out as many contesting explanations as possible through the inclusion of control variables. But

indubitably there are decisive aspects of teaching, such as the level of demand or the speed of teaching, which may influence the students' cognitive involvement, that were not considered within the study.

Second, the two predictor variables are distributed rather unevenly. While most of the courses involved teacher-guided methods – many of them even predominantly, quite a large number of courses did not employ activating methods at all or only in very small doses. On a descriptive level the information about the occurrences of the methods is quite an interesting finding in itself. However, the analyses would certainly become more reliable and more meaningful if the sample contained more courses with larger portions of activating methods.

Third, it is important to note that learning achievement and development of academic competencies were assessed via student self-report. This might affect the validity of the measurement. Self-reported indications of learning outcomes may in particular be influenced by the students' likes and dislikes of the teaching. Nevertheless, as students' self-ratings of their achievement have been shown to correspond with objective achievement measures (e.g., Cohen, 1981), students' subjective assessment of their learning progress was deemed a suitable indicator in this study. Due to strong deviations in grading styles between but also within disciplines, it did not seem appropriate to use the actual grades given by the teachers. Similarly, it was not feasible to construct and use an overall achievement test, since the courses covered completely different content and were on different levels of demand. Concerning the evaluation of competencies, observation of actual behavior or some kind of hands-on test might seem to be more valid modes of assessment at first sight. However, self-report of competencies is still widely accepted as an appropriate measure (cf. Braun et al., 2012). Moreover, the academic competencies investigated in this study, which include, among others, a different mind-set or a change in values and approaches to scientific questions, may prove difficult to assess in a more practical context, as they may manifest differently across disciplines and are hardly observable.

Fourth, the time point of measuring student cognitive involvement – the mediator in our theoretical model – may be a limitation. As it was assessed retrospectively in the end of the semester together with the three final criteria, the associations among these variables might be overestimated.

Fifth, another point of critique may concern the measurement of the predictors: Of course, three observed sessions are only a small fraction of a whole course and there may be doubts as to whether this "extract" really reveals the usual teaching methods used. The reliability of the three ratings is satisfactory, though, indicating a good stability of measurement.

Sixth, we did not consider the varying academic disciplines in our study. It is possible that there are differential effects in certain subjects. Due to the sample size, analyses with subsamples were impossible. However, it was also the primary aim of this study to investigate preferably general relations between teaching methods and student learning, independent of disciplines.

Last, aside from aspects pertaining to the measurement of the investigated constructs, the reader may perceive the lack of a quality check of the employed teaching methods to be a serious limitation of the presented study. Of course, a measure assessing appropriateness and implementation quality of the different teaching methods would have been helpful to determine

whether the reason for their effects lies with implementation issues rather than with the method itself. From our point of view, however, an assessment which is applicable to diverse ways of teaching and allows for their comparison seems hardly feasible. Instructional methods such as lecturing or group work come with specific requirements for a high-quality implementation, so that general quality criteria may not exist.

5.4.3 Conclusion

Even though this study may have its shortcomings and the effects of the distinct teaching methods must not be overrated, the results make an important contribution to the empirical base for educational theory building and political decision making.

Do our findings indicate that university teachers should stop being the guide on the side and return to being the sage on the stage (King, 1993)? We refrain from deducing this kind of prescription. However, the empirical data suggest that there might be a disadvantage in using student-activating methods, whereas teacher-guided learning formats seem to be beneficial. We therefore do call into question the blind plea for activating methods in higher education and stress the need for a stronger empirical basis – and as such, for additional meaningful studies. The results presented cast doubt on the quality of activating methods currently employed in university teaching. Any advances towards increased use of activating methods in higher education would need to be accompanied by concrete recommendations concerning measures of quality assurance.

6 Substudy on the Output Measurement of Higher Education Teaching [9]

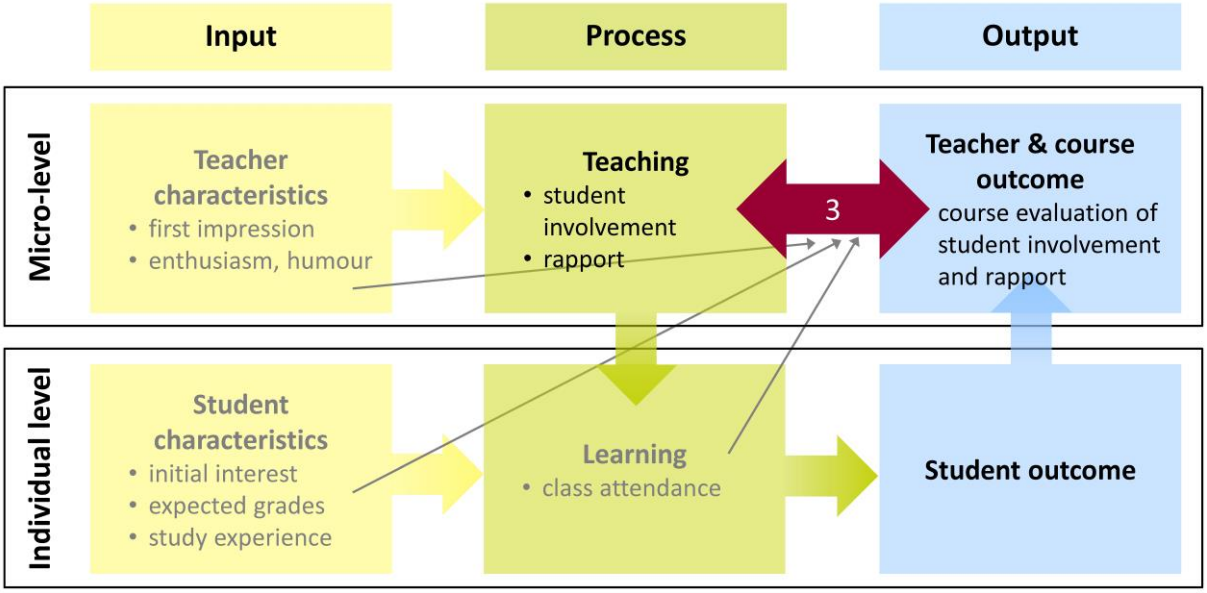

Figure 6.0. Illustration of the third substudy within the framework of higher education teaching.

[9] This chapter was co-authored by Martin Hänze and published in Assessment & Evaluation in Higher Education. Please cite:
Fischer, E., & Hänze, M. (2019). Bias hypothesis under scrutiny: investigating the validity of student assessment of university teaching by means of external observer ratings. *Assessment and Evaluation in Higher Education*, *44*, 772-786. doi.org/10.1080/02602938.2018.1535647 (open access)
Minor modifications (e.g., numbering of tables, figures and headings, alphabetical order of references within parentheses) were made to increase consistency within the dissertation. Sources are still cited in a way that also allows for separate reading (citation of all authors with first mention in chapter).

Bias Hypotheses under Scrutiny: Investigating the Validity of Student Assessment of University Teaching by means of External Observer Ratings

Abstract

To advance the discussion on the validity of student evaluations of university teaching, student ratings of two teaching dimensions – student involvement and rapport – were compared with corresponding observer ratings. Seven potential bias variables were tested with regard to their impact on the students' teaching assessment: three teacher characteristics (first impression, enthusiasm, humour) and four student characteristics (prior interest, expected grades, study experience, class attendance). Bias was defined as an impediment of the students' assessment of teaching on course level. By means of bivariate correlations with course averages and two-level latent moderated structural equations, data of 1,716 students in 80 courses was analysed. Results showed that all three teacher characteristics were genuinely connected to rapport, but even explained variance of the student-rated variable when controlling for observer-rated rapport. The assessment of student involvement was not modified by the teacher characteristics except for teacher enthusiasm, which affected the student evaluation when controlling for observed involvement and, moreover, moderated the relation between the observed and the student-rated variable. For the examined student characteristics, no biasing effects were found – neither on rapport nor on student involvement.

Keywords: higher education; teaching; student evaluation; validity; bias hypothesis

6.1 Introduction

There are few topics that have triggered as much research in higher education research as the validity of student evaluations of university courses. The bulk of studies was already conducted in the 1970s and 1980s, which led to early meta-analyses (e.g., Abrami, Leventhal, & Perry, 1982; Cohen, 1981; Feldman, 1989) and broad agreement about fundamental aspects concerning the reliability and usefulness of student evaluations as a measure to assess higher education teaching. In 1987, Herbert W. Marsh summarized the current literature in a comprehensive review and concluded: "Student ratings are clearly multidimensional, quite reliable, reasonably valid, [and] relatively uncontaminated by many variables often seen as sources of potential bias" (p. 369).

Thirty years later, the validity of student evaluations is still being discussed. Spooren, Brockx, and Mortelmans (2013) reviewed studies published since 2000 and organised them within a validity framework. Among other things, they differentiated between papers examining the convergent validity of student evaluations, e.g., by comparing student ratings to measures of student achievement or to observed teaching behaviour, and papers checking the discriminant validity of student evaluations, i.e., the multitude of bias studies. In alignment with other scholars (such as Stehle, Spinath, & Kadmon, 2012), the authors noticed in the literature a strong focus on the relationship between student ratings and possible biasing factors (Spooren et al., 2013). Marsh (2007) views with some scepticism the recent research on potential biases in student evaluations and complains that it was "frequently atheoretical, methodologically flawed, (…) not based on well-articulated operational definitions of bias, [and] thus continuing to fuel (and to be fuelled by) myths about bias" (p. 346).

This study aimed to advance the research on the validity of student ratings of university teaching by using observational data as an "anchor" to thoroughly revise the effects of a few "classic" bias variables. Doing so, we present a clear definition of what is understood by bias, and use state-of-the-art statistical analyses to check the assumptions. To show how the study fits into the realm of validity research, we first review the relevant strands of existing research.

6.1.1 Convergent and Discriminant Validity of Student Evaluations

The *convergent validity* of a construct refers to the assumption that the method of assessment should be irrelevant, so that two distinct measures of the same construct should correlate substantially (cf. Shadish, Cook, & Campbell, 2002). In the case of student evaluations, which are supposed to assess teaching quality, student learning outcomes are often regarded as the best alternative measure (e.g., Marsh, 1987, p. 286). However, multisection studies, which compare identical courses that deploy the same achievement measure but differ in instructional aspects, delivered inconsistent results as to the agreement between evaluation and test scores (e.g., Clayson, 2009; Cohen, 1981; Uttl, White, & Gonzales, 2017).

Another approach to investigating the convergent validity of student evaluations is to compare them with the judgement of other evaluators – former students, the teachers themselves, colleagues, administrators, or trained observers. Marsh (1987) compiled the results of studies on the agreement of different sources assessing university teaching and reported that faculty self-evaluations and student ratings were usually moderately correlated; evaluations by colleagues or

administrators based on classroom visits, in contrast, repeatedly lacked reliability and produced no substantial correlations to student ratings.

In regard to ratings by external observers as a measure of university teaching, studies of the early research phase deliver quite a clear picture: Murray (1983), for example, conducted a study, in which sets of 18 to 24 trained observers visited classes taught by university lecturers who had previously received low, medium, or high student ratings and documented the occurrence of 60 specific, low-inference teaching behaviours. The observations indeed varied among the three criterion groups of instructors; group differences were particularly noticeable for teaching aspects in the areas of clarity, enthusiasm, and rapport. The link between the observed teaching behaviour and the student ratings of previous courses suggests that student evaluations are determined by actual classroom behaviour of the instructor, which supports their validity. More recently, Renaud and Murray (2005) not only correlated ratings of students and observers, but even compared the pattern of the distinct ratings. The structure of student ratings showed a moderate relation to observed behaviours for low-inference items (e.g., praise students for good ideas); for high-inference items (e.g., foster student participation) no significant association between patterns could be found. By and large, trained observers were determined a valuable source of information for investigating the validity of student evaluations.

Overall, studies assessing the convergent validity of student evaluations were mostly affirmative. The finding that different sources usually agree better when concrete, low-inference criteria were rated concurs with the claim of Marsh and Roche (1997), who reject a narrow, criterion-related approach to assessing the validity of student evaluations, e.g., by measuring performance only. Instead, they request that specific factors of student evaluations be correlated with other specific, supposedly converging teaching variables as an adequate proof of validity.

Discriminant validity describes the extent to which a certain construct is unrelated to other variables it is not theoretically connected to (cf. Shadish et al., 2002). In the context of student evaluations, bias studies seek to demonstrate to what extent the students' ratings are associated with variables that do not pertain to the actual teaching. Usually, they investigate whether so-called "bias variables" lead to a more or less favourable rating of teaching. Many bias hypotheses are based on information-processing models and the notion that student ratings do not only depict instructional effectiveness but also reflect students' cognitive processes while rating the instructors. "Raters use supraordinate features, like general impressions, to attend to, store, retrieve, and integrate judgements of specific behaviors" (d'Apollonia & Abrami, 1997, p. 1200). So, students may rate specific dimensions of instruction on the basis of their global evaluation. According to Spooren and colleagues (2013), this halo effect, i.e., the students' failure to discriminate among conceptually distinct and potentially independent aspects of the teacher's behaviour due to strong general opinions, is a matter still subject to lively discussion.

Reviews of the early studies agreed that biasing variables played a minor role in student ratings of instruction (Marsh, 1987; Marsh & Roche, 1997); recent studies tend to confirm that conclusion (Beran & Violato, 2005; Carle, 2009). Variables that were found to correlate with student ratings are students' prior subject interest, expected and actual grades, reasons for taking a course, workload and difficulty, class size, level of the course, and academic discipline. For most relations,

the effects tended to be small, the directions of the effects were sometimes inconsistent, and the interpretation was not always clear (Marsh, 1987). As Marsh (2007) noted, bias variables may also partly support the validity of student evaluations by demonstrating how they influence the effectiveness of teaching. For instance, class size is negatively correlated with group interaction and rapport but not with other teaching dimensions – a plausible result pattern that properly reflects the way class size impacts student learning.

6.1.2 Research on Potential Bias Variables

The variables that were checked with respect to their biasing effect on student evaluations include circumstantial variables (e.g., weather, time or mode of evaluation), course variables (e.g., format, level, subject discipline, size), teacher variables (e.g., experience, gender, attractiveness, likeability), and student variables (e.g., age, gender, interest, expected grade). Among the teacher characteristics suspected to unduly influence the students' perception of teaching are constructs like teacher charisma (Shevlin, Banyard, Davies, & Griffiths, 2000) and instructor likeability (Clayson & Sheffet, 2006; Delucchi, 2000), his or her expressiveness, humour or entertainment qualities (Abrami et al., 1982; Garner, 2006), and the first impression a teacher gives (Clayson & Sheffet, 2006). On the side of the student characteristics that may influence the assessment of higher education teaching, pre-course interest (Olivares, 2001) and expected grades (Beran & Violato, 2005; Centra, 2003) are the most prominent, but more manifest variables like gender, study experience (Santhanam & Hicks, 2001), and class attendance (Beran & Violato, 2005; Ting, 2000) were also investigated. For a number of classic bias suspects, the research results are reported in more detail:

Instructor expressiveness is one construct that was studied extensively early on. The so-called Dr Fox effect (Naftulin, Ware, & Donnelly, 1973), the overriding influence of instructor expressiveness on students' evaluations of university teaching and the concurring contention that an enthusiastic lecturer can entice favourable evaluations, triggered ample follow-up research: Abrami et al. (1982) conducted a review and meta-analysis of Dr Fox studies and found that expressiveness manipulations had substantial impact on global student ratings and a small effect on achievement, while content quality had a substantial effect on achievement and a small effect on ratings. They stated, however, that "while the summary of prior findings is clear, the implications of the findings for the validity of student ratings is not" (p. 456). d'Apollonia and Abrami (1997), for instance, pondered that expressivity might have a meaningful influence on student ratings as it actually affects student learning. Garner (2006) showed that students in lectures with humorous insertions perceived the lessons as better and the mode of communication and the instructor as more positive; plus, they recalled significantly more learning content afterwards. Wanzer, Frymier, and Irwin (2010) also reported that an instructor's sense of humour was associated with the students' positive affect toward the instructor and the course as well as with their learning behaviours; content-related humour was specifically the most effective.

With regard to *first impressions* made by teachers, various studies showed that personality judgements based on only brief instances, partly only a few seconds and sometimes even without sound, proved significantly related to end-of-term evaluations of teaching (Ambady & Rosenthal, 1993; Clayson & Sheffet, 2006; Tom, Tong, & Hesse, 2010). While Clayson and Sheffet (2006)

interpreted their results as undermining the validity of student evaluations, Tom and colleagues (2010) concluded that personality variables are likely an inherent and inextricably integrated component of instruction, and thus a natural part of student evaluations.

Students' *prior subject interest* was found to be one of the most influential student variables (Marsh, 1987). Notably, Marsh found it to be more highly correlated with perceived learning than with any other dimension of teaching effectiveness. The very same pattern was observed by Olivares (2001). Furthermore, instructor self-evaluations of their teaching were also positively correlated with both their own and their students' perception of student interest (Marsh, 1987). In view of these findings, Marsh argued that student interest might not be a bias of student ratings, but rather influence some aspects of effective teaching.

In many studies, students' *expected grades* have been determined as positively related to ratings of teaching effectiveness. Marsh (1987) presented three different explanations for class-average expected grades to correlate positively with student ratings: 1) the grading leniency hypothesis, suggesting that instructors who give higher-than-deserved grades are rewarded with higher-than-deserved student ratings, which would be serious bias; 2) the validity hypothesis, proposing that better-than-expected grades reflect better student learning triggered by better teaching; and 3) the student characteristic hypothesis, assuming that student variables such as prior interest affect the students' learning process and thus their grade expectancy. Of the 15 papers reviewed by Spooren et al. (2013), 12 report a positive relation between expected grade and student evaluation of teaching, while two find no significant association and one finds differential results that depend on the teaching dimension. In a multiple regression analysis with many course and student variables, for example, Beran and Violato (2005) found the students' expected grades to be the main predictor of course evaluations, but it accounted only for 6% of the variance in the teaching assessment. In his study of about 55,000 courses, Centra (2003) conducted a multiple regression analysis with expected grades and various control variables – e.g., student effort and involvement, class size, class level, course format, institution type – on course evaluation. He concluded that since the effects on course evaluations were so minimal, a bias as suggested by the grading leniency hypothesis did not exist.

Investigations of the role of *class attendance* produced clear findings: Students who attend most classes usually submit higher evaluation scores (e.g., Beran & Violato, 2005). Ting (2000) conducted analyses on course level and equivalently reported that higher class attendance rates (ratio of students present to class size) concurred with more favourable evaluations. However, it is obvious that the direction of this relation may well be backward; that is, students who do not appreciate the teaching in a course will more likely stay away.

Concerning the students' *study experience* and the respective level of the courses they attend, only small effects have so far been found, with advanced courses often receiving slightly higher ratings (e.g., Santhanam & Hicks, 2001). Notably, both student ratings and faculty self-evaluations tend to be higher in graduate courses than in undergraduate courses (Marsh, 1987).

6.1.3 Critique on Bias Research

To Marsh (1987), many studies inspecting the discriminant validity of student evaluations appear to have been conducted "quick and dirty", prompting him to note, "the search for potential biases to student ratings has itself been so biased, that it could be called a witch hunt" (p. 328). Following Marsh (1987, p. 309), important and common methodological problems in the search for potential biases include 1) using correlation to argue for causation; 2) neglecting the distinction between practical and statistical significance; 3) failing to consider the multivariate nature of student ratings; 4) selecting an inappropriate unit of analysis – class average responses nearly always being the appropriate unit of analysis; 5) targeting replicability and generalizability; and 6) lacking an explicit definition of bias against which to evaluate effects.

Two of these methodological issues were taken up explicitly by this study and are therefore described in further detail. First, the proper unit of analysis: A major part of the studies that doubt the validity of student ratings justify the scepticism with associations between evaluation scores and other variables on individual level (Beran & Violato, 2005; Clayson & Sheffet, 2006; Shevlin et al., 2000). Herein lies one of the major pitfalls of evaluation research. Size and even direction of correlations obtained when nested data is analysed on an individual level may differ from correlations based on class-average responses. For student characteristics thought to bias the assessment of teaching, Marsh (1987, p. 281) pointed out that even if some characteristics influenced individual student responses, they would have little effect on class-average responses as long as they were distributed evenly across courses. Thus, the practical significance of biasing variables can hardly be explored on the student level without considering the course affiliations.

With regard to the sixth point of critique, the lack of explicit definitions, Marsh (1987) warned insistently to avoid interpreting the mere existence of a correlation between variables and student evaluation scores as support for bias hypotheses, and called for theoretically defensible delineations of what constitutes a bias. As an ambitious definition, he proposed, for instance, to speak of bias only if student evaluations were influenced by variables that are unrelated to teaching effectiveness and if the impact extended to all dimensions of teaching rather than being specific to particular aspects. d'Apollonia and Abrami (1997) went even further in defining the requirements for a variable to be considered as biasing: "Although many variables have been shown to influence student ratings of instruction, unless they can be shown to moderate the validity coefficient (the correlation between student ratings and student learning), they cannot be described as biasing variables" (p. 1202). Thus, only if a variable led to a weakening of the students' accordance with another criterion should it be considered a bias variable. However, to our knowledge only few studies (e.g., Cohen, 1981) checked for moderating effects of potential bias variables.

6.1.4 Rationale of this Study

To investigate bias in student evaluations, this study built on measures of convergent validity – the assessment of two aspects of teaching rated by students on one hand and by external observers on the other. Assuming that external observers quite objectively evaluate university teaching, as they were explicitly trained to do, their rating was regarded as the gold standard, to which the students' rating had to align in order to be considered valid. The two teaching aspects

examined in this study are both well established and important dimensions of quality teaching (cf. Feldman, 1989; Marsh, 1987; Schneider & Preckel, 2017): active student involvement and rapport.

The corresponding measures of the two teaching aspects allow for two approaches to explore the validity of student evaluations. The first approach follows up on an idea by Marsh (1987, p. 312) and primarily provides insight on genuine links between potential bias variables and teaching dimensions: Potential bias variables are correlated with the student-rated and with the observed teaching variables. A significant relation to the students' measure of teaching is the necessary condition for a biasing effect. If, however, the variable relates to the observed teaching measure as well, an actual link between the two constructs can be inferred. While this approach does not allow for any conclusions with regard to bias, it does provide evidence concerning genuine connections of, for instance, teacher characteristics and certain teaching aspects.

The second approach focuses on the potential bias of evaluation data and is a bit more complex, combining the bias concepts of Marsh (1987, p. 312) and d'Apollonia and Abrami (1997, p. 1202). There are two ways student assessment of teaching can be impaired: First, evaluating university courses students may consistently assign higher or lower values than appropriate; and second, the student assessment may diverge inconsistently and impede the reliability of the evaluation. To determine whether student assessment of teaching is indeed impaired in one or the other way, regression analyses with an observed teaching aspect and a potential bias variable as predictors and the student-rated teaching aspect as criterion are informative. Here, main effects of potential bias variables indicate their undue influence on the students' teaching assessment. So, if a variable shares variance with a student-rated teaching dimension over and above the actual teaching that was captured by the observers, we can infer bias. By way of including interaction terms the analyses can further reveal whether potential bias variables impact the reliability of the student assessment. A negative interaction effect would indicate a weakening of the relation between observed and student-rated teaching dimension through the bias variable. As this study did not seek to research the way a single student generates his or her assessment, but instead whether student evaluations of whole courses are biased, we focused on the course level effects. Only if a potential bias variable had an effect on course values, would it be ascribed a biasing effect.

The delineated approaches to move forward on the question of validity of student evaluations were tested with a range of seven distinct "bias suspects": three teacher characteristic and four student characteristics. The teacher variables were the students' first impression of the teacher and two facets of instructor expressiveness, enthusiasm and humour. Even though these constructs are popular bias suspects, it is still unclear whether or not they are actually related to teaching effectiveness, or if empirical associations to evaluation scores indicate bias. Two student variables that range among the top candidates for bias hypotheses in evaluation research are prior interest and expected grades. Usually these are thought to affect the level of the evaluation scores; it remains unclear, however, whether their effect is of practical significance and whether they might also impact the reliability of the students' ratings. The students' study experience and attendance are two variables that may be associated with the evaluated quality of teaching and that may also enhance the students' ability to reliably assess the teaching they experience. The expertise to adequately evaluate university courses and the ability to differentiate distinct aspects may grow with

study experience, and the knowledge of the concrete teaching to be evaluated will be more accurate with regular attendance. If the subject of judgment is known only superficially, the judgement will also be more strongly influenced by general opinion.

6.2 Method

6.2.1 Sample

The study was conducted at a middle-sized public university in Germany. In preparation, 180 university teachers were asked to take part with the incentive of receiving feedback on their teaching afterwards; the participation rate was 44%. The resulting sample consisted of 79 teachers giving 80 different courses (one teacher used the option of participating twice, with a lecture and a seminar). The courses were 48 lectures and 32 seminars and varied considerably in number of participants (*M* = 59.8, *SD* = 62.8, *Min* = 8, *Max* = 386 students at the beginning of the semester). Lectures often imply large student numbers and teachers imparting a broad subject matter, whereas seminars usually come with small learning groups and an elaboration of more specific learning content with stronger student participation. The courses covered a wide range of disciplines, among others philosophy, foreign languages, economics, sociology, and physics. Of all students enrolled in the courses, 1,716 students took part in the entry survey as well as the final survey and were thus included in this study. They were on average *M* = 23.2 years old (*SD* = 4.3) and had been studying for *M* = 4.2 (*SD* = 2.6) semesters; 59% were female. Participation was voluntary.

6.2.2 Procedure

The data was collected as part of a bigger research project on teaching in higher education during three semesters from 2014 to 2015. Before each semester, the teachers were informed about the general aim and the procedure of the study and chose a course for the subsequent investigation. At the beginning of the semester, the students enrolled in the respective courses filled out an entry questionnaire. During the semester, each course was visited three times by trained observers, who rated various aspects of the teaching on a standardized form. At the end of the semester, prior to final exams, the students completed another questionnaire, in which they rated various teacher and course characteristics.

6.2.3 Instruments

The teaching dimensions student involvement and rapport were assessed by students and observers. The students rated them retrospectively at the end of the semester. *Student-rated student involvement* was measured with five items (α = .88, e.g., “The teacher involved the students actively in the course”, “It was important to the teacher that students thought along and participated”, *M* = 4.4, *SD* = 1.0) and the *student-rated rapport* with seven items (α = .89, e.g., “The teacher is open for questions and problems of the students”, “The teacher meets the students with respect”, *M* = 4.9, *SD* = 0.9) on a 6-point Likert-scale (1 *disagree*, 6 *fully agree*). The agreement of the students within the courses ranged from $ICC_{1,1}$ = .27 for rapport to .37 for student involvement, which is comparable to previous findings in student evaluation data (cf. Marsh, 2007, p. 333). The

reliability of the course mean values, which is of primary significance for analyses on course level, ranged between $ICC_{1,k}$ = .89 and .93. When aggregated course-wise, both variables correlated with r = .55, p < .001.

The observers were student research assistants who had completed approximately 24 hours of training involving the assessment of online lectures as well as joint life observations of university courses and extensive discussions about the single ratings with the first author. They inspected the courses (unannounced) three times, using a standardized rating form to describe each session. About 25% of the visits were done jointly to determine the degree of agreement between raters. The three ratings per course were aggregated to one value, ratings of joint visits being included as an average. The observed teaching characteristics were each measured with one item: *Observed student involvement* ('The students are engaged and participate actively'; M = 2.9, SD = 0.9) and *observed rapport* ('The teacher is respectful, friendly, and appreciative toward the students. He is attentive, open for other opinions and suggestions, takes student questions and comments seriously, and lets them finish'; M = 4.2, SD = 0.5) were both rated on a 5-point Likert scale, with higher values indicating higher agreement. The inter-rater reliabilities were calculated separately for each semester and ranged between $ICC_{1,1}$ = .59 and .75 for student involvement and between $ICC_{1,1}$ = .37 and .49 for rapport. The correlation between the two observed teaching measures was r = .28, p < .05.

The variables whose biasing influence was tested were all reported by the students. The *first impression* of the teacher was assessed in the entry questionnaire with five items asking to what extent the students believed the teacher to be fair, accessible and understanding, well organised, entertaining, and a good instructor (α = .78); answers were given on a 6-point Likert-scale (1 *no, not at all*, 6 *yes, very much so*). *Teacher enthusiasm* and *humour* were assessed in the final questionnaire with seven items (α = .91; e.g., "The teacher is teaching with great enthusiasm") respective two items (α = .87; e.g. "The teacher is humorous") on a 6-point Likert-scale (1 *disagree*, 6 *fully agree*). The inter-rater reliability among students of the same courses was $ICC_{1,1}$ = .22 for first impression, .28 for enthusiasm and .38 for humour; the course averages had a reliability of $ICC_{1,k}$ = .85, .89 and .93 in the same order.

The students' initial interest, their expected grades and study experience were assessed with the entry questionnaire at the start of the semester. *Initial interest* was measured with seven items (α = .89; e.g., "I find many of the topics covered in this course very interesting") on a 6-point Likert-scale (1 *disagree*, 6 *fully agree*), whereas the *expected grades* were assessed with the open question, "With what grade do you think you will pass this course?" As the German grading system uses small numbers for high achievement, the item was inverted so that higher values indicated higher expectations, and marked the expectation of better grades. *Study experience* was captured by the current semester number. Lastly, regular *attendance* was assessed in the final questionnaire by asking the students about the estimated number of sessions they missed during the semester; the item was inverted as well so that higher values indicated more regular attendance. Due to a mistake in a part of the measurement instruments and the resulting missing values, the analyses involving attendance build on data from 70 courses only. The correlations between the student ratings within the courses were $ICC_{1,1}$ = .23 (initial interest), .26 (expected grades), .44 (study experience), and .17

(attendance); the course averages showed an agreement of $ICC_{1,k}$ = .87 (initial interest), .88 (expected grades), .94 (study experience), and .81 (attendance), respectively.

All variables were Z-standardized.

6.3 Results

As this study built on measures of convergent validity, the correspondence between the two differently measured variables of the same teaching dimensions was checked first. The student ratings correlated significantly with the respective observational measures: r = .67, p < .001 for student involvement and r = .48, p < .001 for rapport. Hence, the precondition for the subsequent analyses was met.

For the investigation of genuine links between potential bias variables and teaching dimensions, the student-reported variables were all aggregated on course level to compute bivariate correlations between the teacher and student characteristics on the one side, and the student-reported and the observed teaching dimensions on the other. As can be seen in Table 6.1, the correlation pattern tentatively shows parallels between the potential bias variables and the two distinct measures for student involvement and rapport, respectively. The coefficients of the associations to the student-reported teaching dimensions are mostly higher than the ones to the observed measures, so that the associations with observed student involvement do not always reach significance. The concordant effects suggest that a number of the teacher and student characteristics might not necessarily bias the students' assessment, but rather be genuinely related to the teaching dimensions. Notably, attendance does not relate to either of the student-reported variables, but to the observed teaching characteristics.

Table 6.1

Correlations of the Teaching Measures with Potentially Biasing Teacher and Student Characteristics on Course Level, N = 80

	Student involvement		Rapport	
	Student-rated	Observed	Student-rated	Observed
Teacher characteristics:				
First impression	.28*	.12	.61**	.28*
Enthusiasm	.42**	.17	.78**	.37**
Humour	.31**	.21	.73**	.35**
Student characteristics:				
Initial interest	.26*	.19	.21	.09
Expected grades	.23*	.31**	.21	.10
Study experience	.17	.22*	.09	-.00
Attendance	.03	.24*	.11	.25*

Note. Student-reported variables were aggregated course-wise.
*p < .05. **p < .01.

To determine whether the teacher and student characteristics biased the students' assessment of teaching, two-level latent moderated structural equations (cf. Preacher, Zhang, & Zyphur, 2016) were computed with Mplus (Muthén & Muthén, 1998-2014), using full information maximum likelihood estimation to handle missing values. Preacher and colleagues (2016) advanced statistical methodology to examine different kinds of moderation effects in two-level data more accurately. In the latent structural equation models they recommend, the variances of the student-reported variables are split into their parts within and between the different courses by means of latent factors. Other than in conventional structural equation models with latent factors, these factors do not represent the shared variance of different items, but on course level capture the shared variance of the students within a course and on student level the individual variance. This statistical approach was chosen because the course variance, which is decisive for questions concerning the validity of student evaluations, could be analysed without the individual student variance, making the analysis more appropriate and more exact. As the hypothesized moderation effects ranged on the course level, the interaction terms were calculated with the latent factor of the potentially moderating variables on course level and the respective observed variable. Figure 6.1 shows the general statistical model of the conducted analyses, while Figure 6.2 illustrates one exemplary analysis testing the impact of teacher enthusiasm on the students' assessment of student involvement. An overview of the results of the 14 separate two-level latent moderated structural equations is presented in Table 6.2.

The most relevant coefficients for answering the question of bias are the main effects of the diverse potential bias variables on course level, as well as the respective interaction terms. Significant main effects are to be found for the three teacher characteristics, particularly on rapport. The only moderation effect is indicated by the negative interaction term of enthusiasm on student involvement.

As opposed to interactions obtained in experiments, in observational studies the interaction effects are usually of small magnitude, disordinal interactions being detected more easily than ordinal ones (Cohen, Cohen, West, & Aiken, 2003, p. 298). According to the sensitivity analysis conducted to compute the effect size of detectable effects using the G*Power 3.1 programme by Faul, Erdfelder, Buchner, and Lang (2009), the analyses were fit to show medium-sized effects (f^2 = .17). This one-level computation was used as a measure of approximation due to a lack of respective tools for two-level latent models.

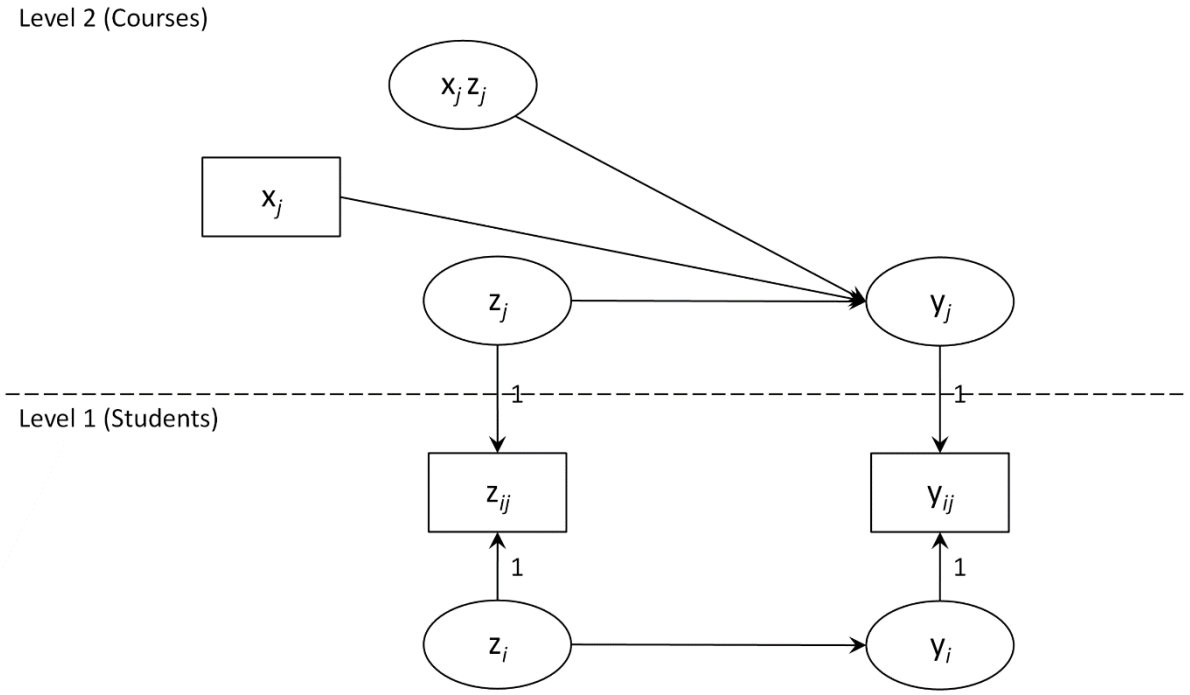

Figure 6.1. Path diagram for the two-level latent moderated structural equation analyses with an interaction term on level 2.

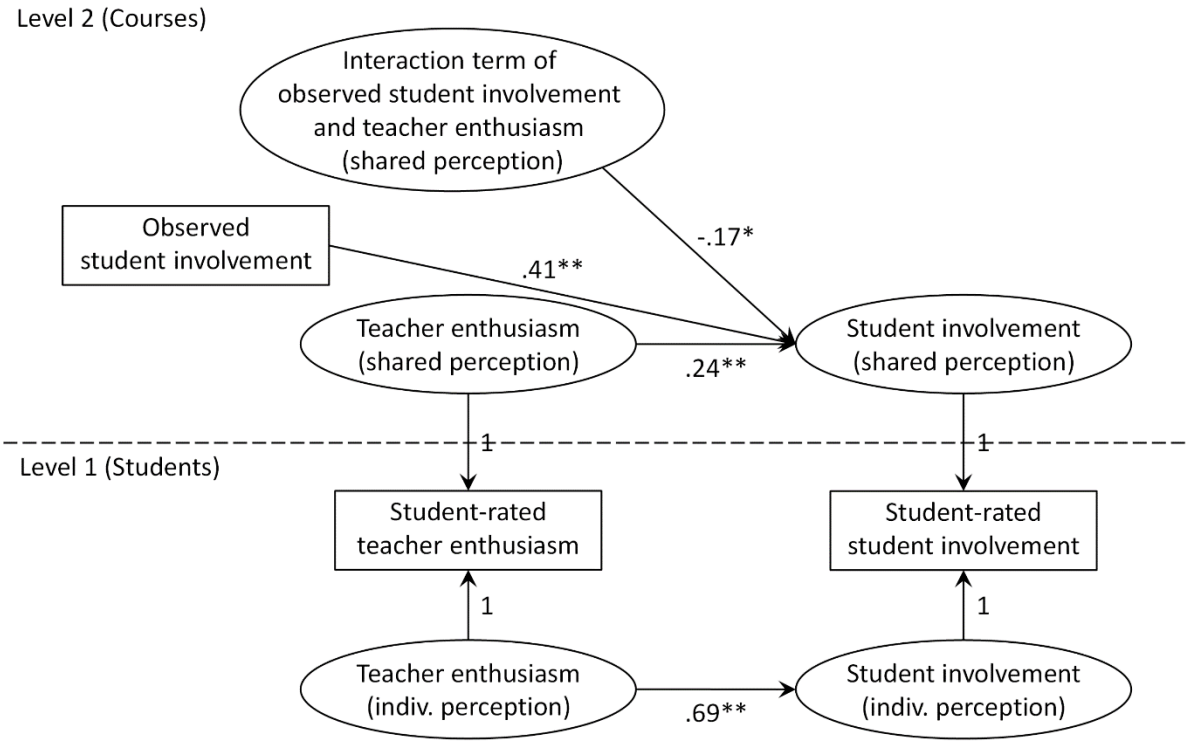

Figure 6.2. Path diagram for the two-level latent moderated structural equation analysis of observed student involvement and student-rated teacher enthusiasm on student-rated student involvement. *p < .05. **p < .01.

Table 6.2

Regression Coefficients on Student Level (L1) and Course Level (L2) of Fourteen Separate Two-level Latent Moderated Structural Equation Analyses

	Student-rated student involvement	Student-rated rapport
Teacher characteristics:		
First impression (L1)	.58**	.60**
Observed student involvement/rapport (L2)	.42**	.20**
First impression (L2)	.20	.60**
Interaction term (L2)	-.06	-.04
Enthusiasm (L1)	.69**	.77**
Observed student involvement/rapport (L2)	.41**	.12**
Enthusiasm (L2)	.24**	.68**
Interaction term (L2)	-.17*	.00
Humour (L1)	.65**	.68**
Observed student involvement/rapport (L2)	.42**	.14**
Humour (L2)	.08	.55**
Interaction term (L2)	-.12	-.03
Student characteristics:		
Initial interest (L1)	.39**	.48**
Observed student involvement/rapport (L2)	.42**	.29**
Initial interest (L2)	.14	.13
Interaction term (L2)	.00	.12
Expected grades (L1)	.16	.22
Observed student involvement/rapport (L2)	.43**	.29**
Expected grades (L2)	.07	.20
Interaction term (L2)	-.06	.21
Study experience (L1)	-.09	.08
Observed student involvement/rapport (L2)	.43**	.28**
Study experience (L2)	.08	.11
Interaction term (L2)	-.05	-.00
Attendance (L1)	.30*	.25*
Observed student involvement/rapport (L2)	.45**	.29**
Attendance (L2)	-.17	.00
Interaction term (L2)	.13	.17

*$p < .05$. **$p < .01$.

6.4 Discussion

Student evaluations of higher education teaching have been a matter of dispute ever since they came into being, among both practitioners and researchers. The notion that students are easily misled and not really capable of accurately judging teaching has been particularly persistent. Despite of the many studies conducted to solve the question of bias, conflicting results have kept the discussion alive. This paper acted on the warnings of leading scholars to investigate the issue openly, with a clear definition of bias and thorough methodological conduct, and checked a number of variables frequently suspected to bias the students' assessment of teaching: the first impression of a teacher as well as enthusiasm and humour on the teachers' side, and initial interest, expected grades, study experience, and class attendance on the students' side. In the following, the results are discussed in detail.

6.4.1 Indication of the Validity of Students' Assessment of Teaching

It is firstly notable that the distinct measures of the two teaching dimensions correlated substantially. Also, the observed teaching aspects were more closely related to the student-reported variables assessing the same aspect than to any other variable. This differential correlational pattern can be seen as a first hint toward the validity of the students' assessment.

The correlational analysis following the first approach further checked whether the investigated teacher and student characteristics were at all related to the student-reported teaching dimensions, and if so, whether they were also associated with the respective observed measure. Apart from study experience and attendance, all potential bias variables were indeed related to the student-rated teaching variables, the teacher characteristics showing particularly high associations with rapport. Correlations of the teacher and student characteristics with the observed teaching dimensions were mostly lower and not always significant. Here again, the teacher characteristics concurred most clearly with observed rapport. The teacher characteristics' link to student involvement was less clear; associations to the student-rated variable were not backed by significant correlations with the observed measure, even if the coefficients tentatively pointed in the same direction. Thus, while the existence of genuine relations between the teacher characteristics and rapport could be established, it remains unclear, whether the students' assessment of student involvement and rapport was additionally biased by the investigated teacher characteristics.

The student characteristics initial interest and expected grades were linked to student involvement, but were not significantly connected to rapport. These findings are plausible as it is likely that courses with a higher level of student interest would enable more interaction; the teachers' rapport, however, should not depend as strongly on the students' preconditions. Remarkably, study experience and attendance showed no link whatsoever to the student-rated teaching aspects. The significant relations with the observed measures were quite straightforward, though. It is likely that courses with a higher average study experience would display greater involvement, and clearly, more student involvement and better rapport might encourage and thus correlate with attendance. All in all, the correlational pattern was plausible and revealed genuine connections between a number of constructs.

Following our second approach, we tested the bias hypotheses with fourteen separate two-level latent moderated structural equation models, so that the variance on course level could be separated from the variance on student level and then be targeted explicitly. By computing moderation analyses, two possible effects of the potential bias variables were checked: 1) the association of the potential bias variables with the student assessment while controlling for the observer rating, and 2) the moderating influence of the potential bias variables on the agreement between the student-rated and the observational teaching assessment. The effect pattern was rather clear-cut: The student characteristics did not show any main effects on the two teaching dimensions on course level at all. The teacher characteristics, in contrast, did partly contribute to the explanation of the students' teaching assessment. Especially student-rated rapport shared variance with the ascribed expressiveness and with the students' first impression over and above the rating of the observers. As stated previously, this effect usually indicates bias. Hence, we would conclude that the students rated the courses with, say, enthusiastic teachers more favourably beyond the actual relation of enthusiasm to the teaching dimensions.

However, for a candid interpretation, statistical issues, which particularly concern variables that are closely related to the criterion variable (see Table 6.1) have to be taken into account. There are two aspects that may have led to an underestimation of the observed variable's effect and consequently an overestimation of the teacher characteristics' effect: The confounding of the two related variables makes them split the criterial variance they can both explain. In doing so, the variable with the better measurement – in this case the student-rated teacher characteristics – can claim more variance (also indicated by the comparably low regression coefficients of observed rapport). The advantage of the student-rated variable over the observed variable that arises from the different measurement quality is further spurred on by the specific model computed: Being assessed on course level the observer variable does not enter the analysis with a latent factor, which would consider its measurement error, while the student-rated bias variable enters the analysis latently. So, while the effect of enthusiasm on student involvement should indeed be considered a bias, the same cannot be affirmatively stated with regard to the main effects on rapport. As enthusiasm is the only teacher characteristic affecting both teaching dimensions likewise, according to Marsh's restrictive definition (1987, p. 312) it would be the only one to be considered a bias at all.

One interaction term, that of enthusiasm on student involvement, was significant: With greater teacher enthusiasm the students' shared perception of student involvement diverged more from the observers' assessment. This is, however, the only indication of a variable leading the students' assessment of teaching astray from the observational measure, and thus fulfilling the requirements of d'Apollonia and Abrami (1997, p. 1202) for a biasing variable. All the other interaction terms were at best of small magnitude and did not reach significance. Even the two student characteristics study experience and attendance, for which a moderating effect would have been plausible, did not show any distorting influence on the students' assessment of involvement and rapport. Thus, none of the distinct student characteristics seem to influence the student assessment of teaching unduly.

Having discussed the effects on course level, which are from our point of view ultimately decisive in the discussion about student course evaluations, we want to now draw attention to one

interesting finding at the student level: Except for two variables, all teacher and student variables have significant and partly very high associations with the teaching assessment on the individual level (cf. Table 6.2). Yes, students who are more interested from the start rate their course more favourably than do those who enter a course disinterested. So, if results were interpreted at the student level without taking the course affiliation into account, a strong bias could be postulated. However, in this sample at least, initial interest is seemingly distributed evenly enough over the investigated courses to not cause a bias on the course level.

6.4.2 Limitations of the Study and Suggestions for Future Research

There are some issues concerning our study that may be viewed critically: First, we are well aware that the axiom of the objective, unbiased observers delivering the gold-standard measurement can be questioned; actually, we cannot say anything definite about the validity of the observer ratings. Even so, seeing the many studies discussing bias in student evaluation without any external criteria to compare to the student data, from our point of view, observations as a parallel measure still present a great advantage. Plus, as the observers received extensive training, aimed explicitly at fostering the objectivity, reliability, and validity of the observational ratings, we find the use of these ratings as anchors for the assessment of teaching justified.

Another point of critique also concerns the observers: The interrater reliability of the ratings of the teaching dimensions was not really satisfactory and indicated some ambiguity among raters, especially for rapport. As previously stated, this limitation hampers the informative value of our analyses, in particular with regard to the interpretation of the effects of the teacher characteristics. Nevertheless, in view of the correlations with the respective student-reported teaching dimensions, the observational variables were still considered acceptable measurements of the two constructs.

We have two recommendations on how to increase interrater reliability between observers: A broader foundation of the instrument, i.e., a number of specific items instead of one global item, should reduce measurement error. A more specific assessment will be more accurate, leave less room for interpretation, and will thus lead to higher reliability and better validity. Apart from that, we recommend to train with exemplary courses that vary substantially in the targeted aspects. That way, the observers will learn better how to use the range of the scale in accordance with each other.

Lastly, the regression coefficients hinted that there might be additional small-sized effects that could possibly be discovered with a larger sample. With respect to the call for practical significance, however, the presented results, which suggest that biasing impacts are hardly meaningful, should suffice.

6.4.3 Conclusion

So, what conclusions can be drawn from the findings? What do they say about the validity of student evaluations? To be sure, the results showed that the teaching dimension of rapport is closely connected to teacher characteristics like enthusiasm and humour, and that a first impression of an instructor already forecasts that end. Beyond these genuine relations, our analyses suggested that the positive teacher characteristics of enthusiasm or humour led to more favourable ratings. It was teacher enthusiasm, too, that seemed to impede the students' assessment of student involvement.

However, no impairing moderation effect was found for the other teacher variables or for the teaching dimension of rapport. Hence, our findings remain inconclusive with regard to the influence of teacher characteristics. Regarding the student characteristics, however, our analyses rendered a clear picture: There was no indication for bias far and wide, even though certain relations could have been expected. Thus, seeing that of the 14 analyses with 28 possible indications for bias, only five effects reached significance, notably with some of the most prominent bias variables and a measurement that favoured the "bias candidates", we conclude that all in all our results support the validity of student assessment of teaching. From our point of view, student evaluations should not be regarded as a faultless measure, but as a very valuable indicator of teaching quality that is not easily outperformed by other ways of assessment and deserves to be taken seriously.

For future research on the validity of student evaluations, we advocate clear definitions of what constitutes bias as well as methodological procedures that allow for the targeted analysis of relevant variance components. We caution in particular about declarations against student evaluations that are based on analyses on the individual level. Only if scholars keep a cool head and treat the matter with the required diligence can we hope to progress on questions of validity regarding student assessment of teaching.

7 General Discussion

This study aspired to make a significant contribution to the research on higher education teaching. To examine teaching and learning as they occur in everyday university life, an observatory, naturalistic design was chosen. With its longitudinal set-up involving several measurement points during one semester, the study covered the teaching and learning occurring in university courses comprehensively. Because variables of the input, the process, and the output dimension of higher education were assessed, it was possible to capture the process character of teaching and learning and to consider diverse factors that impact higher education. Data were contributed by three sources – teachers, students, and external observers. This allowed for a sound measurement of personal characteristics and opinions of teachers and students as well as for a multi-perspective assessment of teaching. The sample comprised full data sets of 79 teachers giving 80 courses with overall 5,765 students, of which 1,716 students took part in both the entry and the closing survey. The study covered a broad range of subject disciplines in order to decrease the probability of finding subject-specific effects only.

Within this study, three disparate aspects were investigated in more detail. Each of them focused on a different dimension of higher education teaching. As the findings and implications of the substudies were already discussed thoroughly in the previous chapters, here, I only want to remind the reader of the central findings to then integrate them within the familiar framework of higher education teaching (Figure 7.1). Afterwards, the concept of teaching quality in higher education is discussed in the light of the three studies, and limitations pertaining to the study as a whole are pointed out. The chapter concludes with remarks on practical implications of research on higher education teaching.

7.1 Synopsis of the Central Findings

The first substudy considered the input dimension of higher education teaching and focused on faculty characteristics and their impact on teaching. It showed that the personal value lecturers attach to the teaching task has a significant impact on multiple teaching aspects, such as the quality of instruction, student involvement, and rapport. In this way, the study enhanced the knowledge on university teachers' motivation and underlined its significance. Moreover, the substudy delivered detailed information on the specific effects that faculty teaching beliefs have on the teaching practice as observed by external raters and students. Thereby, it added a "missing link" to previous research, which mostly relied on self-report measures of teaching (e.g., Norton et al., 2005; Prosser & Trigwell, 2006) or examined relations between teaching beliefs and student variables without considering the actual process of teaching at all (Kember & Gow, 1994; Trigwell et al., 1999). Both teachers' task value and their beliefs about teaching proved to be valuable predictors of higher education teaching. Notably, constructivist beliefs were exclusively related to student-reported variables of teaching. This might point to a general attitude of faculty with higher constructivist teaching beliefs, which might be noticed positively by the students and influence their rating.

The second substudy addressed the process dimension and picked up the long tradition of research on teaching methods (cf. Dunkin & Barnes, 1986; McKeachie, 1990). It investigated the effect that teacher-guided and student-activating teaching methods have on student cognitive involvement, interest, learning achievement, and academic competencies. The results indicated oppositional effects of the two kinds of methods: While the use of teacher-guided methods seemed to promote student learning, the use of student-activating methods tended to have negative effects. The study showed that teacher-guided methods should by no means be thought of as leaving the students somehow passive. At the same time, it raised questions regarding the effectivity of student-activating methods, thereby challenging the call for activating teaching prevalent in research and public (e.g., German Rectors' Conference, 2013; King, 1993; Loyens & Gijbels, 2008; Schaper, 2012). Apart from that, the study endorsed that such disparate learning outcomes as perceived knowledge gain, the development of academic competencies, and interest in the subject matter may all be dependent on students' cognitive involvement. Thus, it may be worthwhile for teachers to put effort into raising and keeping students' attention.

The third substudy took a slightly different approach to researching higher education and focused on the measurement of output indicators of teaching. More precisely, it investigated the validity of student course evaluations. Aiming to advance the extensive body of literature on that matter (d'Apollonia & Abrami, 1997; Marsh, 1987; Spooren et al., 2013), the study presented an explicit definition of bias and applied state-of-the-art statistical methods to check seven teacher and student variables that are frequently suspected to bias the students' assessment of teaching. As an advantage over many other evaluation studies that solely relied on student-reported data (e.g., Abrami et al., 1982; Delucchi, 2000), it used the teaching measures provided by the trained observers for comparison with the students' evaluation. As a primary finding, a genuine relation between the students' first impression of a teacher and his expressiveness on the one side and teachers' rapport with the students on the other was established. Apart from that, is was demonstrated that variables such as the students' initial interest and class attendance may be linked to their personal evaluation of teaching, but that these associations do not necessarily become relevant on course level, that is, they do not automatically impede student course evaluations. With regard to the central issue of inquiry, the close relations between the observed teaching aspects and the respective student-reported variables as well as the few effects that indicated an undue influence or distortion of the student assessment generally supported the validity of the student reports. Specifically, student characteristics did not show any effects on the joint teaching assessment at all. However, the included teacher characteristics, and most consistently enthusiasm, did affect the student rating. Thus, student evaluations may not be impeccable, but all in all present a very valuable indicator of teaching quality.

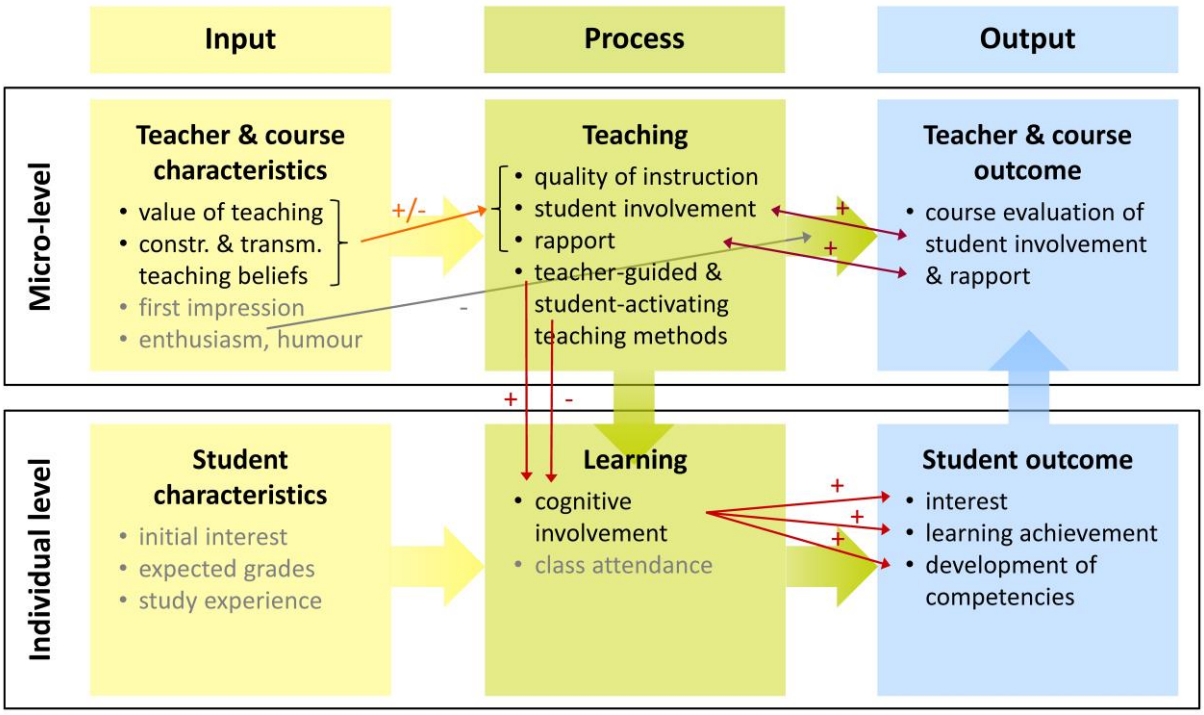

Figure 7.1. Simplified illustration of this study's findings within the framework of higher education teaching.

Taken as a whole, the study delineated the process character of teaching and learning in higher education. The connectedness of the distinct dimensions was particularly highlighted by the first two substudies, but even the third study disclosed how specific input and process factors impact output variables in higher education.

Remarkably, all three studies hinted towards the importance of the university teacher as a person. Evidently, multiple aspects of teaching are influenced by teachers' values and beliefs. And other teacher characteristics, which may be even more decisive such as subject expertise or teaching skills were not even considered. The substudy on teaching methods further suggested that teachers might also be relevant as role models the students learn from. Students may indeed learn best about subject matter and scientific conduct etc. from a scholar relaying certain content and implicitly or explicitly revealing his approach to research in his field. A person with expertise and also enthusiasm may by itself have a positive impact on student learning. It was the teachers' enthusiasm, too, that was shown to have a considerable effect in the third substudy. The teaching of enthusiastic teachers was appraised favourably by the students. Now, while that may be a challenge for the reliable measurement of teaching quality, this result also provides information about the teaching task itself: The potential bias of the student ratings could also be seen as a beneficial process indicator, a supportive aspect of teaching, which might also effect learning through generating a positive attitude to the course and possibly enhancing attention on the students' side. In the end of the day, it should be the factors that are conducive to learning that we should firstly be interested in and strive for, and not the flawless assessment of the teachers' performance.

Lastly, a word on methods: The disparate ways of analysis in the substudies showed that different approaches of dealing with multilevel data are conceivable and may be appropriate. However, it is pivotal that the nesting is considered at all and that the analyses fit the respective research question (cf. in particular Chapter 6). With regard to the usefulness of student reports on higher education teaching, a largely positive conclusion can be drawn. As was indicated by the first and the third substudy, student assessment of teaching may be related systematically to certain teacher characteristics, such as their constructivist teaching beliefs or their enthusiasm. Overall, student assessment of higher education teaching is evaluated as reasonably reliable, though.

7.2 Final Thoughts on Teaching Quality in Higher Education

7.2.1 Descriptive and Normative Approaches to Teaching Quality

Pondering on the basic meaning of quality in the run-up of this study, both the neutrally descriptive and the normative denotations of the term seemed promising with regard to the investigation of higher education teaching. The normative understanding of quality was further delineated using the definition of quality teaching by Fenstermacher and Richardson (2005). However, the definition's component of good teaching in the sense of morally defensible content and methods largely receded into the background as it was presumed to be given. Thus, the primary understanding of quality teaching in a normative sense was strongly connected to its effectiveness. Criteria for success of higher education teaching are manifold and both defined by educational scholars (cf. Shavelson, 2010) and society (cf. Hochschulrahmengesetz, 2017). In this study, for example, students' learning achievement as well as the development of interest and academic competencies were included as desirable learning outcomes and thus indicators for normative quality teaching.

The two different meanings of quality were applied by the first two substudies, respectively: The substudy on the input dimension of higher education teaching impartially explored relations between characteristics of instructors and their way of teaching. The effects on the inspected aspects of teaching were viewed rather neutrally. Strictly speaking, however, the mere choice of variables – both for the input and the process dimension – was already influenced by thoughts on their relevance for student learning, and thus teaching quality in its normative sense. The second substudy embraced the normative perspective on teaching from the outset. Aside from purely describing the prevalence of the two kinds of teaching methods and their respective effects on student learning, it held an evaluative component by explicitly focusing on desirable outcomes as indicators of teaching effectiveness. The distinction between the elements of normative quality teaching (Fenstermacher & Richardson, 2005) became particularly evident and interesting here, as student activating methods are often regarded as good teaching (Duffy et al., 1993; King, 1993; Schaper, 2012; Steffe & Gale, 1995), implicating that the task of teaching is done well. With its somewhat counterintuitive results, this study showed forcefully the challenge that comes with marking process variables as indicators of teaching quality. Furthermore, it underlined the necessity to verify statements on good teaching with measures of student success to eventually accomplish quality teaching.

Aside from the present study, higher education research may use and profit from the different understandings of quality: The descriptive sense of quality may prove particularly valuable for basic research on the functioning of teaching and learning and helps to keep an open mind. However, when higher education research strives to assess the fulfilment of certain standards, the normative perspective is called for. In particular information on how to improve teaching effectivity can only be generated with a clear understanding of the requested outcomes and an evaluative approach towards higher education teaching.

7.2.2 Input, Process, and Output Variables as Indicators of Teaching Quality

To conclude, I want to sketch a few thoughts concerning the use of input, process, and output variables as indicators of quality teaching in its normative sense. These reflections focus on the micro-level of higher education, as it is the level that the process of teaching is situated on. Traditionally, the German higher education system has sought to establish a high level of quality by regulating certain input factors (Braun et al., 2014) with a focus on teachers' subject expertise. With the close connection between research and teaching in faculty positions, teachers are meant to be experts in their respective fields and as such well suited for conveying their subject matter to students, thereby delivering high-standard teaching. This logic of quality assurance may not always apply as the overlap between course contents and the teachers' field of research is sometimes small (cf. Chapter 3.2). A profound expertise in every aspect of teaching content may certainly be beneficial, but a lack of specific knowledge may not hamper the quality of teaching either. An even more important consequence of tying research to teaching is the resulting high probability that teachers are enthusiastic about their subject. As we have learnt, this input aspect may be quite decisive for quality teaching. However, enthusiasm for the own research is one matter, enthusiasm for teaching quite another, and the two do not always cohere. As this study showed, the teachers' personal task value of teaching as well as their beliefs about teaching also significantly influence teaching. And whereas content knowledge and knowledge about teaching can be enhanced (Penny & Coe, 2004; Stes et al., 2010), these fundamental characteristics of faculty, which largely determine the actual way of teaching, may hardly or only laboriously be altered with training programmes (cf. Ebert-May et al., 2011). All in all, this study confirmed the general importance of input factors for quality teaching. Yet, the central input factor currently taken into account in the selection processes for university teachers in Germany is subject expertise indicated by research-related achievements, while motivational and attitudinal aspects with regard to the teaching task play a subordinate role (Kleimann & Hückstädt, 2018). So, in short, quality of higher education teaching can be assessed using input factors; the possibilities to use this dimension to regulate quality of higher education might not be exhausted in Germany, though.

The process dimension of higher education, that is, the actual teaching and learning, constitutes the core of research on teaching (Dunkin & Barnes, 1986). However, as the second substudy has vividly demonstrated, the classification of teaching aspects as indicators of quality is not always straight forward and may produce disagreement among scholars. As with the input factors, in the end, any process variable needs to be legitimated by empirical evidence on its effect on desirable student outcomes. Despite the challenges this dimension poses concerning the

identification of proper quality criteria, it may be viewed as the most important one, as it affects student learning most directly. Reliable information on the effectiveness of different aspects of higher education teaching is required for any deliberate improvement of teaching. Teachers and teacher trainers firstly rely on a sound appraisal of process variables in order to know how to teach effectively. Certainly, the body of knowledge here – largely stemming from investigations using student evaluations – is considerable, providing numerous criteria of quality teaching within the process. At the same time, many questions pertaining to the impact of certain teaching aspects on student learning still need answering. In any case, in Germany, a better reception of the existing empirical findings by the professional teacher trainers seems desirable. In contrast to international literature (e.g., Fry et al., 2009; Perry & Smart, 2007), too many German guidebooks are written without much reference to the empirical foundations of the presented recommendations (e.g., Böss-Ostendorf & Senft, 2014). So yes, in sum, process variables are indicative when assessing the quality of teaching, and several criteria have already been established by international research. They are particularly relevant for efforts pertaining to the improvement of teaching.

Regarding the output dimension of higher education teaching, Donabedian's statement (1966/2005, p. 694) of the student outcomes being the ultimate validators for any teaching seems valid. Aspects of the other two dimensions always have to be justified by the output they produce, if higher education systems are to be effective. The broad, statutory aims of higher education determined by the public are refined by the faculty administrating the single degree programmes. Within these boundaries, the teachers individually define the learning outcomes they pursue in their distinct courses. Thus, the teachers themselves define the immediate output criteria for their instruction, the content and level of the corresponding examination, and even the rigor in grading. As a result, measures of student learning achievement are usually not comparable between courses. Hence, it is hardly possible to use objective student output data for the assessment of teaching quality. Therefore, subjective measures need to be used to tap teaching quality in the output dimension. As confirmed by this study and many others (e.g., Marsh, 2007) student assessment of teaching largely provides valid indication on teaching. However, student course evaluations mostly report on teaching, but cannot capture the actual student learning outcomes in a way that allows for comparison either. So, while the students' learning outcomes are the most important criterion for the teaching quality in a course, they are very difficult to assess. The output dimension of university teaching represents the ultimate measurement point for quality teaching; the measurement itself remains to pose challenges, though.

7.3 General Limitations of the Study

Many significant limitations have already been mentioned in the previous chapters, which investigated specific aspects of higher education teaching. Which limitations concern the whole project and should be considered when conducting a follow-up research with similar layout?

Some challenges and restrictions come with the basic design of the study and simply are the downside of aspects that may otherwise represent particular strengths of the study: For example, definite causal inferences are excluded from the outset, due to the observational nature of the study

and the resulting correlational design. Similarly, the inclusion of two measurements of student data, which increases the validity of the assessment of preconditions and final thoughts on a course, respectively, usually entails a significant loss of sample size when combining the data. Apart from that, the use of multiple data sources – a clear asset of this study – and, specifically, the disparate ways of measurement further result in dissimilar measurement errors, since the reliability of variables varies strongly. As could be seen in the third substudy in particular, this can complicate the joint analyses of data from different sources. Lastly, by delimiting the study on lectures and seminars, other and possibly more innovative formats were excluded ex ante. So, even though this may be a small limitation only, it is still important to note that this study – by far – did not tap the whole range of possible teaching forms and formats.

Aside from the limitations that come with the fundamental out-set of the study, there are also limitations that stem from the implementation. A central weak spot of this study was the partially low interrater reliability of the observers. The inconsistent measurement impacted the resulting effect sizes and impeded the informative value of the studies. Recommendations with respect to this problem were given in the discussion of the third study. They pertained to the instrument – using several specific items instead of one global item to measure a construct (see the Appendix for the modified rating form) – and to the training of the observers – selection of courses for practice that vary specifically in the targeted aspects.

Another weakness concerns the measurement of the students' learning activities. To avoid inaccuracy induced by the retrospective assessment, an intermediate survey inquiring the students' actual learning behaviour during the semester would have been favourable. Unfortunately, this was not feasible here, as the two other student surveys already used up a lot of lecturing time, and the demand of even more class time might have encountered resistance from the teachers. At the worst, this could have jeopardised the sample.

As mentioned before, the assessment of learning achievement via student self-reports can be criticised. However, as it has been pointed out at various places within this study, an objective measure of learning outcomes that is valid for courses covering different content is virtually impossible (cf. Marsh, 1987, p. 287). Thus, the procedure chosen in this study was deemed to be the best possible option.

The sample size further restrained the explanatory power of this study. This applies especially to the teacher sample, as a larger number of students would mainly have improved the reliability of their measurements, which was already quite satisfactory; plus, in a naturalistic setting the numbers of students enrolled in courses cannot be augmented arbitrarily. With more participating teachers the analyses could have been more sensitive for small effects on the one side and more accurate by considering more control variables on the other. In addition, a greater number of seminars might especially have been valuable for the substudy on the teaching methods, as it might have increased the amount of student-activating methods and thus provided a broader basis for the analyses. A higher number of teachers and courses in the various disciplines would further have allowed for a better consideration of the impact of the disciplines on the processes of teaching and learning, possibly even for the inspection of moderating effects.

As for any empirical study, the generalisability of this study is limited per se. However, the university the study was conducted at and, more importantly, the educational processes captured within this study are comparable to many other settings of higher education in Germany and even worldwide. In fact, Jacob and Teichler (2011) found out that faculty in Germany is quite similar to faculty in other countries in a variety of aspects. As participation in this study was voluntary, the sample may not represent the full range of university teachers. A reduced variance of teacher characteristics, however, would only imply a probable underestimation of teacher effects rather than overestimation. On the side of the students, no strong selection effect should be assumed, as all students attending the courses were surveyed. Yet, capturing the present students only, the student sample at the end of the semester showed more favourable characteristics, such as a slightly higher degree of conscientiousness. This fact should not impair the validity of the students' reports, though. It not permissible to claim the generalisability of the detected effects across disciplines. Even though the likelihood of discovering general effects may have been enhanced by deliberately including different disciplines, it could not be confirmed. As mentioned above, the single subgroups were too small to test the invariance of effects between distinct disciplines. However, the fact that effects were significant for such a diverse group of courses increases the probability for the effects to be common ones.

For future research, it can be recommended to use observations to investigate teaching. From the author's point of view, they offer various advantages: Course observation by student research assistants, for example in three separate sessions without the instructor knowing the exact dates, presents a relatively unobtrusive and at the same time well-feasible way to assess teaching, other than, e.g., video studies. Furthermore, the data source is independent of teacher or student data, minimizing the possibility of spurious correlations between teaching behaviours and other measures caused by halo effects or implicit personal theories (cf. Murray, 2007). To successfully deploy standardised observational ratings, a thorough training is crucial, and the more high-inference ratings are to be done, the better the training must be. Also, it is certainly beneficial to select senior students as raters, who have considerable experience with university teaching and know how to act discreetly and politely. All in all, the results of this study present a strong argument for multiperspective research in higher education, and the observation of teaching by trained raters certainly poses a valuable option.

7.4 Concluding Remarks

So, does this study have an overall take home message? If so, then maybe it is this one: University teachers matter. Somewhat surprisingly, this key phrase came up as a common theme in the synopsis of the central results. All three substudies at least in part hint towards the significance of faculty – in many ways. Teachers' personal approach to teaching, including their individual motivation and beliefs as well as their standing as persons with expertise and enthusiasm for what they do, were shown to be important. Apart from these aspects, there are several characteristics that may be even more decisive, such as the subject expertise or certain teaching skills that were not even considered in this study. So, the impact that teachers with the entirety of their abilities, attitudes,

and personality facets have on student learning should not be underestimated. Of course, the effects that were found between input, process, and output variables on course level are mostly small in size. But in view of the great freedom in studying that higher education offers and of the paramount influence of the students' own characteristics on learning and achievement, this would also be expected. However, this study suggests that, albeit it may be a small one, teachers do make a difference. Thus, they should strive to make the best of it.

So, can any implications be derived for practice? As was pondered previously, altering faculty's intrinsic task value and their personal theories may be difficult. Possibly, the teaching motivation might be augmented by a pronounced appreciation of the task by the public, the institution, but most importantly within the scholarly community. Support and training of higher education teachers with regard to their teaching can certainly be fruitful, e.g., towards an effective use of diverse teaching methods. Nevertheless, a selection procedure for academics that takes into account characteristics that are relevant for quality teaching may be even more important. If teaching really is to be a central and not merely a subordinate task of faculty, then their respective suitability needs to be valued as highly as their research portfolio.

Regarding the research-practice link in faculty training in Germany, Kröber and Szczyrba (2011, p. 71) remarked that in spite of the growing number of posts and jobs to enhance quality teaching in universities, it is still not clear what the trainers, that is, higher education didactics, actually have to know and on what basis they can operate adequately. Currently, they have the most diverse disciplinary backgrounds, entailing equally diverse knowledge and, possibly, approaches towards university teaching. Despite that, standards for professionals working in higher education didactics are not available thus far, with the discussion still going on. So, yes, sound research on higher education teaching and learning is necessary, but just as indispensable is its reception by the practitioners.

No, this study did neither deliver guidelines for how to improve university teaching that would be suitable for teacher training, nor did it knit a theory of teaching and learning in higher education, which some scholars call for (e.g., Marsh, 2007, p. 374). But a number of relevant issues were picked up to advance knowledge and discussion on higher education teaching in certain respects. Aside from the various findings of this study, the framework of teaching and learning in higher education on the micro- and the individual level will possibly be found useful by other scholars, too. Here, it proved itself with regard to the integration of disparate research results; maybe it can also help towards the construction of a comprehensive theory that may enhance the understanding of teaching quality in higher education and guide both practice and research in the future.

Postscript

Even though I sought to present my work as a rounded off entirety here, this reflects my internal perception only in part. Yes, the three substudies nicely cover the input, process, and output dimensions of higher education teaching, adopt disparate perspectives on quality, and even apply different analytical approaches to deal with multilevel data. But the project does not feel finished quite yet. There is just so much more to investigate and publish on the basis of the collected data! For that reason, to me the dissertation has something of a slightly arbitrary intermediate step, which is necessary due to formal regulations, but also a little random. Possibly, this ambiguous sensation is brought about by the cumulative format of this thesis. I did not set out to compose an opus embracing my whole research project. Instead, I selected aspects for individual publications, that I found interesting myself and that I deemed to be of interest to the scholarly community as well. (Yes, I know, this procedure does not really conform to the Open Science principles.) Maybe I would feel differently, possibly a little more satisfied, had I handed in the thesis with three more article-chapters, delineating the topic of teaching in higher education more comprehensively. But then again, limitations have their advantages, and maybe this constraint really was for the better.[10]

Be that as it may – there are more articles in the planning: Next, I would like to examine the interaction between lecturer and students during class time. More specifically, I would like to analyse the relations between the number and kind of the questions asked by the teacher and the quantity and quality of student replies. Possibly, these associations are moderated by the teachers' rapport, the atmosphere in the course, or class size. Another issue I find exciting is the question of how to raise and keep attention in university courses. Unfortunately, work on the effectiveness of various rhetorical measures has not yielded meaningful results. However, there are also organisational teaching measures that may increase cognitive involvement, such as homework or a particular kind of examination deployed in the end of the semester. Here, student characteristics like conscientiousness, initial interest, or extrinsic motivation with regard to the grade might moderate the relations.

Let's see what becomes of these intentions. Eventually, the feeling of closure on the topic of teaching quality in higher education will set in. And then, surely, the next adventure in educational research will be waiting.

[10] Despite my lamenting, I would not have had it any other way than the cumulative approach! From my point of view, it poses a very good opportunity – still under the guidance and "protection" of the supervisor – to explore the realm of scientific work and find out whether the world of research and peer-reviewed publishing is an attractive career option in the long run.

References

Abrami, P. C., d'Apollonia, S., & Rosenfield, S. (2007). The dimensionality of student ratings of instruction: What we know and what we do not. In R. P. Perry & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 385-456). Dordrecht, Netherlands: Springer.

Abrami, P. C., Leventhal, L., & Perry, R. P. (1982). Educational Seduction. *Review of Educational Research, 52*(3), 446-464.

Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of Personality and Social Psychology, 64*(3), 431-441.

Assen, J. H. E., Meijers, F., Otting, H., & Poell, R. F. (2016). Explaining discrepancies between teacher beliefs and teacher interventions in a problem-based learning environment: A mixed-methods study. *Teaching and Teacher Education, 60*, 12-23.

Atkinson, J. W. (1957). Motivational determinants of risk-taking behaviour. *Psychological Review, 64*(6), 359-372.

Auferkorte-Michaelis, N., Ladwig, A., & Stahr, I. (Eds.). (2010). *Hochschuldidaktik für die Lehrpraxis. Interaktion und Innovation für Studium und Lehre an der Hochschule* [Higher education didactics for teaching practice. Interactions and innovations for study and teaching in higher education]. Opladen, Germany/ Farmington Hills, MI: Budrich UniPress.

Baeten, M., Struyven, K., & Dochy, F. (2013). Student-centred teaching methods: Can they optimise students' approaches to learning in professional higher education? *Studies in Educational Evaluation, 39*(1), 14-22.

Baleghizadeh, S., & Shakouri, M. (2017). Investigating the relationship between teaching styles and teacher self-efficacy among some Iranian ESP university instructors. *Innovations in Education and Teaching International, 54*(4), 394-402.

Bandura, A. (1977). *Social learning theory.* Oxford, England: Prentice-Hall.

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York, NY: Freeman.

Beran, T., & Violato, C. (2005). Ratings of university teacher instruction: How much do student and course characteristics really matter? *Assessment & Evaluation in Higher Education, 30*, 593–601.

Biehler, R., Hänze, M., Hochmuth, R., Becher, S., Fischer, E., Püschl, J., & Schreiber, S. (2013). *Lehrinnovation in der Studieneingangsphase „Mathematik im Lehramtsstudium" – Hochschuldidaktische Grundlagen, Implementierung und Evaluation: Gesamtabschlussbericht des BMBF-Projekts LIMA* [Teaching innovations in the entry phase „Mathematics for teacher students" – Didactic foundations, implementation and evaluation: Final report of the BMBF-project LIMA]. Hannover, Germany: TIB. Retrieved from https://doi.org/10.2314/GBV:791017176 (28.11.2018)

Biggs, J. (2012). Enhancing learning trough constructive alignment. In J. R. Kirby & M. J. Lawson (Eds.), *Enhancing the quality of learning: Dispositions, instruction, and learning processes* (pp. 117-136). New York, NY: Cambridge University Press.

Blüthmann, I. (2012). Individuelle und studienbezogene Einflussfaktoren auf die Zufriedenheit von Bachelorstudierenden [Individual and study-related factors influencing the satisfaction of Bachelor students]. *Zeitschrift für Erziehungswissenschaft, 15*, 273-303.

Bologna Declaration (1999). *The European Higher Education Area. Joint Declaration of the European Ministers of Education*, Bologna, 19 June 1999. Retrieved from http://www.ehea.info/media.ehea.info/file/Ministerial_conferences/02/8/1999_Bologna_Declaration_English_553028.pdf (27.11.2018)

Böss-Ostendorf, A., & Senft, H. (2014). *Einführung in die Hochschul-Lehre. Ein Didaktik-Coach* [Introduction into higher education teaching. A didactic coach] (2nd rev. ed.). Opladen, Germany/ Toronto, Canada: Barbara Budrich.

Braun, E., & Hannover, B. (2011). Gelegenheiten zum Kompetenzerwerb in der universitären Lehre. Zusammenhänge zwischen den Einschätzungen Studierender und unabhängigen Beobachtungen relevanter Merkmale universitärer Lehrveranstaltungen [Opportunities to gain competencies in higher education: Correlation between students' estimations and independent observations of learning opportunities]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 43*(1), 22-28.

Braun, E., & Leidner, B. (2009). Academic course evaluation: Theoretical and empirical distinctions between self-rated gain in competences and satisfaction with teaching behavior. *European Psychologist, 14*(4), 297-306.

Braun, E., Weiß, T., & Seidel, T. (2014). Lernumwelten in der Hochschule [Learning environments in higher education]. In T. Seidel & A. Krapp (Eds.), *Pädagogische Psychologie* (6th rev. ed., pp. 433-453). Weinheim/ Basel, Germany: Beltz.

Braun, E., Woodley, A., Richardson, J. T. E., & Leidner, B. (2012). Self-rated competences questionnaires from a design perspective. *Educational Research Review, 7*(1), 1-18.

Bronfenbrenner, U. (1994). Ecological Models of Human Development. In T. Husén & T. N. Postlethwaite (Eds.), *The International Encyclopedia of Education* (2nd ed., Vol. 3, pp. 1643-1647). Oxford, England: Pergamon.

Brown, J. S., Collins, A., & Duiguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32-42.

Busato, V. V., Prins, F. J., Elshout, J. J., & Hamaker, C. (2000). Intellectual ability, learning style, personality, achievement motivation and academic success of psychology students in higher education. *Personality and Individual Differences, 29*, 1057-1068.

Carle, A. C. (2009). Evaluating college students' evaluations of a professor's teaching effectiveness across time and instruction mode (online vs. face-to-face) using a multilevel growth modeling approach. *Computers & Education, 53*, 429-435.

Centra, J. A. (2003). Will teachers receive higher student evaluations by giving higher grades and less course work? *Research in Higher Education, 44*, 495-518.

Cherney, I. D. (2008). The effects of active learning on students' memories for course content. *Active Learning in Higher Education, 9*(2), 152-171.

Clayson, D. E., & Sheffet, M. J. (2006). Personality and the student evaluation of teaching. *Journal of Marketing Education, 28(*2), 149-160.

Clayson, D. E. (2009). Student evaluation of teaching: Are they related to what students learn? A meta-analysis and review of the literature. *Journal of Marketing Education, 31*(1), 16-30.

Coffey, M., & Gibbs, G. (2002). Measuring teachers' repertoire of teaching methods. *Assessment & Evaluation in Higher Education, 27*(4), 383-390.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/ correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Cohen, P. A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Review of Educational Research, 51*(3), 281-309.

Cole, J. S., Bergin, D. A., & Whittaker, T. A. (2008). Predicting student achievement for low stakes tests with effort and task value. *Contemporary Educational Psychology, 33*, 609-624.

Credé, M., Roch, S. G., & Kieszczynka, U. M. (2010). Class attendance in college: A meta-analytic review of the relationship of class attendance with grades and student characteristics. *Review of Educational Research, 80*(2), 272-295.

Cretchley, P. C., Edwards, S. L., O'Shea, P., Sheards, J., Hurst, J., & Brookes, W. (2014). Research and/or learning and teaching: A study of Australian professors' priorities, beliefs and behaviours. *Higher Education Research & Development, 33*(4), 649-669.

d'Apollonia, S., & Abrami, P. C. (1997). Navigating student ratings of instruction. *American Psychologist, 52*(11), 1198-1208.

Daumiller, M., Grassinger, R., Dickhäuser, O., & Dresel, M. (2016). Structure and relationships of university instructors' achievement goals. *Frontiers in Psychology, 7*, 375.

Delucchi, M. (2000). Don't worry, be happy: Instructor likability, student perceptions of learning, and teacher ratings in upper-level sociology courses. *Teaching Sociology, 28*, 220-231.

Deslauriers, L., Schelew, E., & Wiemann, C. (2011). Improved learning in a large-enrollment physics class. *Science, 332*(6031), 862-864.

Dietrich, J., Viljaranta, J., Moeller, J., & Kracke, B. (2017). Situational expectancies and task values: Associations with students' effort. *Learning and Instruction, 47*, 53-64.

Dochy, F., Segers, M., Van den Bossche, P., & Gijbels, D. (2003). Effects of problem-based learning: A meta-analysis. *Learning and Instruction, 13*(5), 533-568.

Donabedian, A. (2005). Evaluating the quality of medical care. *The Milbank Quaterly, 83*(4), 691-729. (Reprinted from *The Milbank Memorial Fund Quaterly, 44*(3), 1966, 166-203.)

Dresel, M., & H. Rindermann (2011). Counseling university instructors based on student evaluations of their teaching effectiveness: A multilevel test of its effectiveness under consideration of bias and unfairness variables. *Research in Higher Education, 52*, 717-737.

Duffy, T. M., Lowyck, J., & Jonassen, D. H. (Eds.). (1993). *Designing environments for constructive learning.* Berlin, Germany: Springer.

Dunkin, M. J., & Barnes, J. (1986). Research on teaching in higher education. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 754-777). New York, NY: Macmillan.

Ebert-May, D., Derting, T. L., Hodder, J., Momsen, J. L., Long, T. M., & Jardaleza, S. E. (2011). What we say is not what we do: Effective evaluation of faculty professional development programs. *BioScience, 61*, 550-558.

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology, 53*, 109-132.

EHEA Ministerial Conference. (2009). *Leuven-Communiqué: The Bologna Process 2020: The European Higher Education Area in the new decade.* Retrieved from http://media.ehea.info/file/2009_Leuven_Louvain-la-Neuve/06/1/Leuven_Louvain-la-Neuve_Communique_April_2009_595061.pdf (06.04.2017)

EHEA Ministerial Conference. (2012). *Bucharest Communiqué: Making the most of our potential: Consolidating the European Higher Education Area.* Retrieved from http://media.ehea.info/file/2012_Bucharest/67/3/Bucharest_Communique_2012_610673.pdf (06.04.2017)

European Students' Union. (2015). *Overview on student-centred learning in higher education in Europe: Research study.* Brussels, Belgium: European Students' Union. Retrieved from http://media.ehea.info/file/Student_centered_learning/63/2/Overview-on-Student-Centered-Learning-in-Higher-Education-in-Europe_679632.pdf (06.04.2017)

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149-1160.

Feldman, K. A. (1989). Association between student ratings of specific instructional dimensions and student achievement: Refining and extending the synthesis of data from multisection validity studies. *Research in Higher Education, 30*, 583-645.

Fenstermacher, G. D., & Richardson, V. (2005). On making determinations of quality in teaching. *Teachers College Record, 107*(1), 186-213.

Fischer, E. (2018). *Skalendokumentation – Qualität von Hoschullehre* [Scale Documentation – Quality of Higher Education Teaching]. Unpublished manuscript, Department of Psychology, University of Kassel, Germany.

Fischer, E., & Hänze, M. (2019). Bias hypothesis under scrutiny: Investigating the validity of student assessment of university teaching by means of external observer ratings. *Assessment and Evaluation in Higher Education, 44*, 772-786.

Floud, R. (2006). The Bologna process: Transforming European higher education. *Change, 38*(4), 8-15.

Fredricks, J., Filsecker, M., & Lawson, M. (Eds.). (2016). Student engagement and learning: Theoretical and methodological advances [Special issue]. *Learning and Instruction, 43*.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences of the United States of America (PNAS), 111*(23), 8410-8415.

Fry, H., Ketteridge, S., & Marshall, S. (2009). *A Handbook for Teaching and Learning in Higher Education. Enhancing Academic Practice* (3rd ed.). New York, NY: Routledge.

Gargallo López, B., Almerich Cerveró, G., Suárez Rodríguez, J. M., García Félix, E., & Garfella Esteban, P. R. (2013). Learning styles and approaches to learning in excellent and average first-year university students. *European Journal of Psychology of Education, 28*(4), 1361-1379.

Garner, R. L. (2006). Humor in pedagogy. How ha-ha can lead to aha! *College Teaching, 54*(1), 177-180.

Gasiewski, J. A., Eagan, M. K., Garcia, G. A., Hurtado, S., & Chang, M. J. (2012). From gatekeeping to engagement: A multicontextual, mixed method study of student academic engagement in introductory STEM courses. *Research in Higher Education, 53*(2), 229-261.

Geiser, C. (2011). *Datenanalyse mit Mplus. Eine anwendungsorientierte Einführung* (2nd ed.). Wiesbaden: VS Verlag für Sozialwissenschaften.

German Council of Science and Humanities. (2007). *Empfehlungen zu einer lehrorientierten Reform der Personalstruktur an Universitäten* [Recommendations on a teaching oriented reform of the staff structure at universities]. Berlin, Germany: German Council of Science and Humanities. Retrieved from https://www.wissenschaftsrat.de/download/archiv/7721-07.pdf (30.11.2018)

German Council of Science and Humanities. (2008). *Empfehlungen zur Qualitätsverbesserung von Lehre und Studium* [Recommendations on quality improvement in higher education]. Berlin, Germany: German Council of Science and Humanities. Retrieved from https://www.wissenschaftsrat.de/download/archiv/8639-08.pdf (06.04.2017)

German Council of Science and Humanities. (2012). *Prüfungsnoten an Hochschulen im Prüfungsjahr 2010. Arbeitsbericht mit einem Wissenschaftspolitischen Kommentar des Wissenschaftsrates* [Examination grades in higher education in 2010. Work report with a research policy commentary from the German Council of Science and Humanities]. Hamburg, Germany: German Council of Science and Humanities. Retrieved from https://www.wissenschaftsrat.de/download/archiv/2627-12.pdf (20.11.2018)

German Rectors' Conference. (2013). *Recommendation of the 15th general meeting of the German Rectors' Conference (HRK) on 19 November 2013 in Karlsruhe. European study reform.* Bonn, Germany: German Rectors' Conference. Retrieved from https://www.hrk.de/uploads/tx_szconvention/HRK_MV_15_Empfehlung_Europaeische_Studienreform_EN_01.pdf (06.04.2017)

Gonzáles, A., Rodríguez, Y., Faílde, J. M., & Carrera, M. V. (2016). Anxiety in the statistics class: Structural relations with self-concept, intrinsic value, and engagement in two samples of undergraduates. *Learning and Individual Differences, 45*, 214-221.

Grant, A. M. (2008). The significance of task significance: Job performance effects, relational mechanisms, and boundary conditions. *Journal of Applied Psychology, 93*(1), 108-124.

Green, S. B. (1991). How many subjects does it take to do a regression analysis? *Multivariate Behavioral Research, 26*(3), 499-510.

Griffiths, S. (2009). Teaching and learning in small groups. In H. Fry, S. Kettridge, & S. Marshall (Eds.), *A handbook for teaching and learning in higher education. Enhancing academic practice* (3rd ed., pp. 58-71). New York, NY: Routledge.

Han, J., Yin, H., & Wang, W. (2016). The effect of tertiary teachers' goal orientations for teaching on their commitment: The mediating role of teacher engagement. *Educational Psychology, 36*(3), 526-547.

Hancock, D. R., Bray, M., & Nason, S. A. (2002). Influencing university students' achievement and motivation in a technology course. *The Journal of Educational Research, 95*(6), 365-372.

Hänze, M., Fischer, E., Schreiber, S., Biehler, R., & Hochmuth, R. (2013). Innovationen in der Hochschullehre: Empirische Überprüfung eines Studienprogramms zur Verbesserung von vorlesungsbegleitenden Übungsgruppen in der Mathematik [Innovations in higher education: Empirical evaluation of a program to improve tutorials accompanying a mathematics lecture]. *Zeitschrift für Hochschulentwicklung, 8*(4), 89-103.

Hativa, N., & Birenbaum, M. (2000). Who prefers what? Disciplinary differences in students' preferred approaches to teaching and learning styles. *Research in Higher Education, 41*(2), 209-236.

Hattie, J., & Marsh, H. W. (1996). The relationship between research and teaching: A meta-analysis. *Review of Educational Research, 66*(4), 507-542.

Heise, E., Hasselhorn, M., & Hager, W. (2003). Lehrevaluation, Lehrveranstaltungszufriedenheit und Leistung [Evaluation of university teaching, student achievement and satisfaction with courses]. *Psychologie in Erziehung und Unterricht, 50*(1), 43-57.

Helmke, A., Rindermann, H., & Schrader, F.-W. (2008). Wirkfaktoren akademischer Leistungen in Schule und Hochschule [Factors affecting academic achievement in school and university]. In W. Schneider & M. Hasselhorn (Eds.), *Handbuch der Pädagogischen Psychologie* (pp. 145-155). Göttingen, Germany: Hogrefe.

Helmke, A., & Schrader, F.-W. (2010). Hochschuldidaktik [Higher education didactics]. In D. H. Rost (Ed.), *Handwörterbuch Pädagogische Psychologie* (4th rev. ed., pp. 273-279). Weinheim/ Basel, Germany: Beltz.

Hochschulrahmengesetz (1976/1999/2017). Retrieved from https://www.gesetze-im-internet.de/hrg/HRG.pdf (03.12.2018)

Hopback, A. (Ed.) (2004). *Qualitätssicherung an Hochschulen. Neue Herausforderungen nach der Berlin-Konferenz* [Quality assurance in higher education insitutions. New challenges after the Berlin conference]. German Rectors' Conference. Bielefeld, Germany: Bertelsmann.

Hox, J. J. (2010). *Multilevel analysis. Techniques and applications* (2nd ed.). Hove, England: Routledge.

Jacob, A.-K. & Teichler, U. (2011). *Der Wandel des Hochschullehrerberufs im internationalen Vergleich: Ergebnisse einer Befragung in den Jahren 2007/08* [The Change of the Academic Profession in International Comparison: Results of a Survey Conducted in 2007/08]. Bonn, Germany: Bundesministerium für Bildung und Forschung.

Jahnke, I., & Wildt, J. (2011). Editorial: Hochschuldidaktische Hochschulforschung – fachbezogen und fachübergreifend?! [Editorial: Research on higher education didactics – discipline-specific or not?!] In I. Jahnke & J. Wildt (Eds.), *Fachbezogene und fachübergreifende Hochschuldidaktik* (pp. 9-18). Bielefeld, Germany: Bertelsmann.

Jang, H., Kim, E. J., & Reeve, J. (2016). Why students become more engaged or more disengaged during the semester: A self-determination theory dual-process model. *Learning and Instruction, 43*, 27-38.

Johansen, K., Jung, K., Lexa, S., & Niekrenz, Y. (2010). *Einsteigerhandbuch Hochschullehre. Aus der Praxis für die Praxis* [Handbook for beginners in higher education teaching. From practice for practice] (2nd ed.). Darmstadt, Germany: WBG.

Jones, B. D., Paretti, M. C., Hein, S. F., & Knott, T. W. (2010). An analysis of motivation constructs with first-year engineering students: Relationships among expectancies, values, achievement, and career plans. *Journal of Engineering Education, 99*(4), 319-336.

Kane, R., Sandretto, S., & Heath, C. (2002). Telling half the story: A critical review of research on the teaching beliefs and practices of university academics. *Review of Educational Research, 72*(2), 177-228.

Kember, D. (1997). A reconceptualization of the research into university academics' conceptions of teaching. *Learning and Instruction, 7*(3), 255-275.

Kember, D. (2000). Lecturers' approaches to teaching and their relationship to conceptions of good teaching. *Instructional Science, 28*(5-6), 469-490.

Kember, D., & Gow, L. (1994). Orientations to teaching and their effect on the quality of student learning. *The Journal of Higher Education, 65*(1), 58-74.

Kim, T., & Schallert, D. L. (2014). Mediating effects of teacher enthusiasm and peer enthusiasm on students' interest in the college classroom. *Contemporary Educational Psychology, 39*(2), 134-144.

King, A. (1993). From sage on the stage to guide on the side. *College Teaching, 41*(1), 30-35.

Klassen, R. M., & Tze, V. M. C. (2014). Teachers' self-efficacy, personality, and teaching effectiveness: A meta-analysis. *Educational Research Review, 12*, 59-76.

Kleimann, B., & Hückstädt, M. (2018). Auswahlkriterien in Berufungsverfahren: Universitäten und Fachhochschulen im Vergleich [Selection criteria in appointment procedures: Universities and universities of applied sciences in comparison]. *Beiträge zur Hochschulforschung, 40*(2), 20-46.

Kobayashi, K. (2005). What limits the encoding effect of note-taking? A meta-analytic examination. *Contemporary Educational Psychology, 30,* 242-262.

Kröber, E., & Szczyrba, B. (2011). Zwischen disziplinärer Herkunft und hochschuldidaktischer Identität – Auf dem Weg zu professionellen Standards in der Hochschuldidaktik [Between disciplinary heritage and identity as higher education didactics – On the way to professional standards in higher education didactics]. In I. Jahnke & J. Wildt (Eds.), *Fachbezogene und fachübergreifende Hochschuldidaktik* (pp. 69-79). Bielefeld, Germany: Bertelsmann.

Krumboltz, J. D., & Farquhar, W. W. (1957). The effect of three teaching methods on achievement and motivational and outcomes in a how-to-study course. *Psychological Monographs: General and Applied, 71*(14), 1-26.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the graduate record examinations: Implications for graduate student selection and performance. *Psychological Bulletin, 127*(1), 162-181.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology, 86*, 148-161.

Lawless, C., & Richardson, J. T. E. (2004). Monitoring the experiences of graduates in distance education. *Studies in Higher Education, 29*(3), 353-374.

Lee, H. H., Kim, G. M. L., & Chan, L. L. (2015). Good teaching: What matters to university students. *Asia Pacific Journal of Education, 35*(1), 98-110.

Lim, J., Kim, M., Chen, S. S., & Ryder, C. E. (2008). An empirical investigation of student achievement and satisfaction in different learning environments. *Journal of Instructional Psychology, 35*(2), 113-119.

Loveland, J. L. (2014). *Traditional lecture versus an activity approach for teaching statistics: A comparison of outcomes* (Doctoral dissertation). Utah State University, Logan, UT. Retrieved from http://digitalcommons.usu.edu/cgi/viewcontent.cgi?article=3089&context=etd (13.02.2017)

Loyens, S. M. M., & Gijbels, D. (2008). Understanding the effects of constructivist learning environments: Introducing a multi-directional approach. *Instructional Science, 36*(5-6), 351-357.

Lübeck, D. (2009). *Lehransätze in der Hochschullehre* [Approaches to teaching in higher education]. (Doctoral dissertation). Freie Universität Berlin, Berlin, Germany. Retrieved from www.diss.fu-berlin.de/diss/receive/FUDISS_thesis_000000011078 (30.04.2012)

Lüdtke, O., Marsh, H. W., Robitzsch, A., & Trautwein, U. (2011). A 2x2 taxonomy of multilevel latent contextual model: Accuracy-bias trade-offs in full and partial error correction models. *Psychological Methods, 16*, 444-467.

Lüdtke, O., Marsh, H.W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychological Methods, 13*, 203-229.

Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research, 11*, 253-388.

Marsh, H. W. (2007). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases and usefulness. In R. P. Perry & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 319-383). Dordrecht, Netherlands: Springer.

Marsh, H. W., & Roche, L. A. (1997) Making students' evaluations of teaching effectiveness effective. *American Psychologist, 52*(11), 1187–1197.

Marton, F., & Saljö, R. (1976). On qualitative differences in learning – I: Outcomes and processes. *British Journal of Educational Psychology, 46*, 4-11.

Martin, E., Prosser, M., Trigwell, K., Ramsden, P., & Benjamin, J. (2000). What university teachers teach and how they teach it. *Instructional Science, 28*, 387-412.

Mayer, R. E. (2009). Constructivism as a theory of learning versus constructivism as a prescription for instruction. In S. Tobias & T. M. Duffy (Eds.), *Constructivist instruction: Success or failure?* (pp. 184-200). New York, NY: Routledge.

McKeachie, W. J. (1990). Research on college teaching: The historical background. *Journal of Educational Psychology, 82*(2), 189-200.

Mesa, V., Celis, S., & Lande, E. (2014). Teaching approaches of community college mathematics faculty: Do they relate to classroom practices? *American Educational Research Journal, 51*(1), 117-151.

Metz-Göckel, S., Kamphans, M., & Scholkmann, A. (2012). Hochschuldidaktische Forschung zur Lehrqualität und Lernwirksamkeit. Ein Rückblick, Überblick und Ausblick [Didactic research on teaching quality and teaching efficacy in higher education – Retrospective, insights and outlook]. *Zeitschrift für Erziehungsiwssenschaft, 15*, 213-232.

Meyer, J. H., & Eley, M. G. (2006). The Approaches to Teaching Inventory: A critique of its development and applicability. *British Journal of Educational Psychology, 76*, 633-649.

Mitchell, T. R., & Albright, D. W. (1972). Expectancy theory predictions of the satisfaction, effort, performance, and retention of naval aviation officers. *Organizational Behavior and Human Performances, 8*, 1-20.

Morton, A. (2009). Lecturing to large groups. In H. Fry, S. Kettridge & S. Marshall (Eds.), *A handbook for teaching and learning in higher education. Enhancing academic practice* (3rd ed., pp. 58-71). New York, NY: Routledge.

Moses, I. (1985). High quality teaching in a university: Identification and description. *Studies in Higher Education, 10*(3), 301-313.

Murray, H. G. (1983). Low-inference classroom teaching behaviors and student ratings of college teaching effectiveness. *Journal of Educational Psychology, 75*(1), 138-149.

Murray, H. G. (2007). Low-inference teaching behaviours and college teaching effectiveness: Recent developments and controversies. In R. P. Perry & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 145-200). Dordrecht, Netherlands: Springer.

Muthén, L. K., & Muthén, B. O. (1998-2014). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Naftulin, D. H., Ware, J. E., & Donnelly, F. A. (1973). The Doctor Fox lecture: A paradigm of educational seduction. *Journal of Medical Education, 48*, 630-635.

Nezlek, J. B., Schröder-Abé, M., & Schütz, A. (2006). Mehrebenenanalysen in der psychologischen Forschung. Vorteile und Möglichkeiten der Mehrebenenmodellierung mit Zufallskoeffizienten [Multilevel analyses in psychological research. Advantages and potential of multilevel random coefficient modeling]. *Psychologische Rundschau,57*(4), 213-223.

Neumann, R. (1996). Researching the teaching-research nexus: A critical review. *Australian Journal of Education, 40*(1), 5-18.

Neumann, R., Parry, S., & Becher, T. (2002). Teaching and learning in their disciplinary contexts: A conceptual analysis. *Studies in Higher Education, 27*(4), 405-417.

Norton, L., Richardson, J. T. E., Hartley, J., Newstead, S., & Mayes, J. (2005). Teachers' beliefs and intentions concerning teaching in higher education. *Higher Education, 50*, 537-571.

Olivares, O. J. (2001). Student interest, grading leniency, and teacher ratings: A conceptual analysis. *Contemporary Educational Psychology, 26*, 382-399.

Quality (2018). In *Oxford English Dictionary Online.* Retrieved from http://www.oed.com/viewdictionaryentry/Entry/155878 (12.11.2018)

Pant, H. A., & Zlatkin-Troitschanskaia, O. (2016). Valid assessment of student competencies in higher education. Methodological innovations and perspectives for educational measurement Special issue]. *Journal of Educational Measurement, 53*(3).

Parpala, A., Lindblom-Ylänne, S., & Rytkönen, H. (2011). Students' conceptions of good teaching in three different disciplines. *Assessment and Evaluation in Higher Education, 36*(5), 549-563.

Patrick, B. C., Hisley, J., & Kempler, T. (2000). "What's everybody so excited about?": The effects of teacher enthusiasm on student intrinsic motivation and vitality. *Journal of Experimental Education, 68*(3), 217-236.

Penny, A. R., & Coe, R. (2004). Effectiveness of consultation on student ratings feedback: A meta-analysis. *Review of Educational Research, 74*(2), 215-253.

Perez, T., Cromley, J. G., & Kaplan, A. (2014). The role of identity development, values, and costs in college STEM retention. *Journal of Educational Psychology, 106*(1), 315-329.

Perry, R. P., & Smart J. C. (Eds.). (2007). *The scholarship of teaching and learning in higher education: An evidence-based perspective*. Dordrecht, Netherlands: Springer.

Piaget, J. (1952). *The origins of intelligence in children* (M. Cook, Trans.). New York, NY: International Universities Press.

Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review, 16*, 385-407.

Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2016). Multilevel structural equation models for assessing moderation within and across levels of analysis. *Psychological Methods, 21*(2), 189-205.

Pritchard, R. D., & Sanders, M. S. (1973). The influence of valence, instrumentality, and expectancy on effort and performance. *Journal of Applied Psychology, 57*(1), 55-60.

Prosser, M., & Trigwell, K. (2006). Confirmatory factor analysis of the Approaches to Teaching Inventory. *British Journal of Educational Psychology, 76*, 405-419.

Renaud, R. D., & Murray, H.G. (2005). Factorial validity of student ratings of instruction. *Research in Higher Education, 46*(8), 929-953.

Renkl, A. (2009). Why constructivists should not talk about constructivist learning environments: A commentary on Loyens and Gijbels (2008). *Instructional Science, 37*(5), 495-498.

Resnick, L. (1987). Learning in school and out. *Educational Researcher, 16*(9), 13-20.

Richardson, J. T. E. (2005). Students' approaches to learning and teachers' approaches to teaching in higher education. *Educational Psychology, 25*(6), 673-680.

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin, 138*(2), 353-387.

Risko, E. F., Andersin, N., Sarwal, A., Engelhardt, M., & Kingstone, A. (2012). Everyday attention: Variation in mind wandering and memory in a lecture. *Applied Cognitive Psychology, 26*, 234-242.

Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin, 130*(2), 261-288.

Rudinger, G., Krahn, B., & Rietz, C. (Eds.). (2008). Evaluation und Qualitätssicherung von Forschung und Lehre im Bologna-Prozess [Evaluation and quality assurance of research and teaching in the Bologna process]. *Applied Research in Psychology and Evaluation, Vol. 1*. Göttingen, Germany: V&R unipress.

Samuelowicz, K., & Bain, J. D. (1992). Conceptions of teaching held by academic teachers. *Higher Education, 24*, 93-111.

Santhanam, E., & Hicks, O. (2001). Disciplinary, gender and course year influences on student perceptions of teaching: Explorations and implications. *Teaching in Higher Education, 76*, 17-31.

Schaper, N. (2012). *Fachgutachten zur Kompetenzorientierung in Studium und Lehre* [Expertise on competence orientation in higher education]. Retrieved from https://www.hrk-nexus.de/fileadmin/redaktion/hrk-nexus/07-Downloads/07-02-Publikationen/fachgutachten_kompetenzorientierung.pdf (06.04.2017)

Schiefele, U. (1991). Interest, learning, and motivation. *Educational Psychologist, 16*(3), 299-323.

Schiefele, U., Streblow, L., Ermgassen, U., & Moschner, B. (2003). Lernmotivation und Lernstrategien als Bedingungen der Studienleistung. Ergebnisse einer Längsschnittstudie [The influence of learning motivation and learning strategies on college achievement: Results of a longitudinal analysis]. *Zeitschrift für Pädagogische Psychologie, 17*(3/4), 185-198.

Schmidt, H. G., van der Molen, H. T., te Winkel, W. W. R., & Wijnen, W. H. F. W. (2009). Constructivist, Problem-Based Learning does work: a meta-analysis of curricular comparisons involving a single medical school. *Educational Psychologist, 44*(4), 227-249.

Schmohr, M., Müller, K., & Philipp, J. (Eds.) (2018). *Gelingende Lehre: erkennen, entwickeln, etablieren. Beiträge der Jahrestagung der Deutschen Gesellschaft für Hochschuldidaktik (dghd) 2016* [Successful teaching: Recognizing, developing, establishing. Contributions to the annual conference of the German Association for Educational and Academic Staff Development in Higher Education (dghd) in 2016]. Bielefeld, Germany: WBV.

Schneider, M., & Preckel, F. (2017). Variables associated with achievement in higher education: A systematic review of meta-analyses. *Psychological Bulletin, 143*, 565-600.

Schrader, F.-W., & Helmke, A. (2000). Wirksamkeit des Hochschulunterrichts aus Sicht der Studierenden. Eine empirische Studie [Effectiveness of higher education from the students' perspective. An empirical study]. In A. Helmke, W. Hornstein, & E. Terhart (Eds.), *Qualität und Qualitätssicherung im Bildungsbereich. Schule, Sozialpädagogik, Hochschule* (pp. 261-276). Weinheim, Germany: Beltz.

Schwab, D. P., Olian-Gottlieb, J. D., & Heneman, H. G., III (1979). Between-subjects expectancy theory research: A statistical review of studies predicting effort and performance. *Psychological Bulletin, 86*(1), 139-147.

Seidel, T., & Hoppert, A. (2011). Merkmale von Lehre an der Hochschule. Ergebnisse zur Gestaltung von Hochschulseminaren mittels Videoanalysen [Characteristics of teaching at university. Results on characteristics of university seminars using video analyses]. *Unterrichtswissenschaft, 39*(2), 154-172.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Belmont, CA: Wadsworth Cengage Learning.

Shavelson, R. J. (2010). *Measuring college learning responsibly: Accountability in a new era.* Stanford, CA: Stanford University Press.

Shevlin, M., Banyard, P., Davies, M. D., & Griffiths, M. (2000): The validity of student evaluation of teaching in higher education: Love me, love my lectures? *Assessment and Evaluation in Higher Education, 25*(4), 397-405.

Sinatra, G. M., Heddy, B. C., & Lombardi, D. (Eds.). (2015). Engagement in the context of science learning [Special issue]. *Educational Psychologist, 50*(1).

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis. An Introduction to basic and advanced multilevel modeling* (2nd ed.). London: Sage.

Spinath, B., & Brünken, R. (2016). *Pädagogische Psychologie – Diagnostik, Evaluation und Beratung* [Educational Psychology – Diagnostic, Evaluation, and Consultation]. Göttingen, Germany: Hogrefe.

Spinath, B., Hasselhorn, M., Artelt, C., Köller, O., Möller, J., & Brünken, R. (2012). Gesellschaftliche Herausforderungen für Bildungsforschung und -praxis. Beiträge der Pädagogischen Psychologie [Social challenges for educational research and practice. Contributions of educational psychology]. *Psychologische Rundschau, 63*(2), 92-110.

Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research, 83*(4), 598-642.

Springer, L., Stanne, M. E., & Donovan, S. S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of Educational Research, 69*(1), 21-51.

Standing Conference of the Ministers of Education and Cultural Affairs. (2008). *Wettbewerb exzellente Lehre. Eine gemeinsame Initiative der Kultusministerkonferenz und des Stifterverbandes für die deutsche Wissenschaft. Beschluss der Kultusministerkonferenz vom 06.03.2008* [Competition for excellent teaching. A joint initiative of the Standing Conference of the Ministers of Education and Cultural Affairs and the Stifterverband. Resolution on the 06.03.2008] Retrieved from https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2008/2008_03_06-Wettbewerb-exzellente-Lehre.pdf (24.10.2018)

Statistisches Bundesamt (2018a). *Press release 460 / 2018-11-27. Winter semester 2018/2019: More students than ever before enrolled at German institutions of higher education.* Retrieved from https://www.destatis.de/EN/PressServices/Press/pr/2018/11/PE18_460_213.html (16.12.2018)

Statistisches Bundesamt (2018b). *Bildungsfinanzbericht 2018. Ausgaben für Bildung. Tabellenteil* [Report on finances for education 2018. Expenses for education. Tables]. Statistisches Bundesamt. Retrieved from https://www.destatis.de/DE/Publikationen/Thematisch/BildungForschungKultur/BildungKulturFinanzen/BildungsfinanzberichtTabellenteil.html (16.12.2018)

Steffe, L. P., & Gale, J. (Eds.). (1995). *Constructivism in education.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Stehle, S., Spinath, B., & Kadmon, M. (2012). Measuring teaching effectiveness: Correspondence between students' evaluations of teaching and different measures of student learning. *Research in Higher Education, 53*, 888-904.

Stes, A., Coertjens, L., & Van Petegem, P. (2013). Instructional development in higher education: Impact on teachers' teaching behaviour as perceived by students. *Instructional Science, 41*, 1103-1126.

Stes, A., Min-Leliveld, M., Gijbels, D., & Van Petegem, P. (2010). The impact of instructional development in higher education: The state-of-the-art of the research. *Educational Research Review, 5*, 25-49.

Struyven, K., Dochy, F., & Janssens, S. (2008). Students' likes and dislikes regarding student-activating and lecture-based educational settings: Consequences for students' perceptions of the learning environment, student learning and performance. *European Journal of Psychology of Education, 23*(3), 285-317.

Sun, S., Vancouver, J. B., & Weinhardt, J. M. (2014). Goal choices and planning: Distinct expectancy and value effects in two goal processes. *Organizational Behavior and Human Decision Processes, 125*, 220-233.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive Load Theory.* New York, NY: Springer.

Times Higher Education (2018). *Europe Teaching Rankings 2018.* Retrieved from https://www.timeshighereducation.com/rankings/europe-teaching/2018 (02.11.2018)

Ting, K. (2000). Cross-level effects of class characteristics on perceptions of teaching quality. *Journal of Educational Psychology, 92*(4), 818-825.

Tobias, S. (2009). An eclectic appraisal of the success or failure of constructivist instruction. In S. Tobias & T. M. Duffy (Eds.), *Constructivist instruction: Success or failure?* (pp. 335-350). New York, NY: Routledge.

Tobias, S., & Duffy, T. M. (2009). The success or failure of constructivist instruction: An introduction. In S. Tobias & T. M. Duffy (Eds.), *Constructivist instruction: Success or failure?* (pp. 3-10). New York, NY: Routledge.

Tom, G., Tong, S. T., & Hesse, C. (2010). Thick slice and thin slice teaching evaluations. *Social Psychology of Education, 13*(1), 129-136.

Trapmann, S., Hell, B., Weigand, S., & Schuler, H. (2007). Die Validität von Schulnoten zur Vorhersage des Studienerfolgs – eine Metaanalyse [The validity of school grades for academic achievement – a meta-analysis]. *Zeitschrift für Pädagogische Psychologie, 21*(1), 11-27.

Trigwell, K., & Prosser, M. (1996). Congruence between intention and strategy in university science teachers' approaches to teaching. *Higher Education, 32*, 77-87.

Trigwell, K., Prosser, M., & Waterhouse, F. (1999). Relations between teachers' approaches to teaching and students' approaches to learning. *Higher Education, 37*, 57-70.

Turney, J. R. (1974). Activity outcome expectancies and intrinsic activity values as predictors of several motivation indexes for technical-professionals. *Organizational Behavior and Human Performance, 11*, 65-82.

Tynjälä, P. (1999). Towards expert knowledge? A comparison between a constructivist and a traditional learning environment in the university. *International Journal of Educational Research, 31*(5), 357-442.

Ulrich, I. (2016). *Gute Lehre in der Hochschule. Praxistipps zur Planung und Gestaltung von Lehrveranstaltungen* [Quality teaching in higher education. Practical recommendations for planning and conducting university courses]. Wiesbaden, Germany: Springer.

Umbach, P. D. (2007). Faculty cultures and college teaching. In R. P. Perry & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 263-317). Dordrecht, Netherlands: Springer.

Uttl, B., White, C. A., & Gonzales, D. W. (2017). Meta-analysis of faculty's teaching effectiveness: student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation, 54*, 22-42.

Van Eerde, W., & Thierry, H. (1996). Vroom's expectancy models and work-related criteria: A meta-analysis. *Journal of Applied Psychology, 81*(5), 575-586.

Van Rossum, E. J., & Schenk, S. M. (1984). The relationship between learning conception, study strategy and learning outcome. *British Journal of Educational Psychology, 54*, 73-83.

Veiga-Simão, A.-M., Flores, M.-A., Barros, A., Fernandes, S., & Mesquita, D. (2015). Perceptions of university teachers about teaching and the quality of pedagogy in higher education: A study in Portugal. *Journal for the Study of Education and Development, 38*(1), 102-143.

Vermunt, J. D. (2003). The power of learning environments and the quality of student learning. In De Corte, E., Verschaffel, L., Entwistle, N., & Van Merriënboer, J. (Eds.), *Powerful Learning Environments: Unravelling Basic Components and Dimensions* (pp. 109-124). Amsterdam, Netherlands: Pergamon.

Vroom, V. H. (1964). *Work and motivation.* New York, NY: Wiley & Sons.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Cambridge, MA: Harvard University Press.

Wanzer, M. B., Frymier, A. B., & Irwin, J. (2010). An explanation of the relationship between instructor humor and student learning: Instructional humor processing theory. *Communication Education, 59*(1), 1-18.

Webb, C., & Baird, J. H. (1968). Learning differences resulting from teacher- and student-centered teaching methods. *Journal of Higher Education, 39*(8), 456-460.

Wibbecke, G., Wibbecke, A., Kahmann, J., & Kadmon, M. (2016). Lehrenden- und studierendenzentrierte Lehre messen. Ein Beobachtungsbogen für Lehrveranstaltungen an Hochschulen [Measuring teacher-centred and student-centred teaching – Introduction of a classsroom teaching observation sheet]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 48*(4), 184-194.

Wigfield, A., & Eccles, J. S. (1992). The development of achievement task values: A theoretical analysis. *Developmental Review, 12*, 265-310.

Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology, 25*, 68-81.

Wildt, J. (2013). Entwicklung und Potentiale der Hochschuldidaktik [Development and potential of higher education didcatics]. In M. Heiner & J. Wildt (Eds.), *Professionalisierung der Lehre. Perspektiven formeller und informeller Entwicklung von Lehrkompetenz im Kontext Hochschulbildung* (pp. 27-57). Bielefeld, Germany: Bertelsmann.

Wilson, K., Lizzio, A., & Ramsden, P. (1997). The development, validation and application of the Course Experience Questionnaire. *Studies in Higher Education, 22*(1), 33-53.

Winteler, A. (2006). Lehrende an Hochschulen [Teachers in higher education]. In A. Krapp & B. Weidenmann (Eds.), *Pädagogische Psychologie* (5th rev. ed., pp. 334-347). Weinheim, Germany/ Basel, Switzerland: Beltz.

Winteler, A. (2011). *Professionell lehren und lernen. Ein Praxisbuch* [Professional teaching and learning. A practice guide] (4th ed.). Darmstadt, Germany: WBG.

Wosnitza, M., Helker, K., & Lohbeck, L. (2014). Teaching goals of early career university teachers in Germany. *International Journal of Educational Research, 65*, 90-103.

Zee, M., & Koomen, H. M. Y. (2016). Teacher self-efficacy and its effects on classroom processes, student academic adjustment, and teacher well-being: A synthesis of 40 years of research. *Review of Educational Research, 86*(4), 981-1051.

Zumbrunn, S., McKim, C., Buhs, E., & Hawley, L. R. (2014). Support, belonging, motivation, and engagement in the college classroom: A mixed method study. *Instructional Science, 42*(5), 661-684.

Danksagung

Ich habe es als großes Privileg empfunden, jahrelang mit viel Freiheit an einem selbst gewählten Thema forschen zu können. Daher gebührt mein Dank zunächst dem Steuerzahler, der diese Arbeit ermöglicht hat und dem ich letztlich auch Rechenschaft schulde. Ich würde mich freuen, wenn meine Gehaltszahlungen als gut investiertes Geld befunden würden.

M. H. – Martin Hänze, mein Mentor und Helfer. ;-) Was ich dir nicht alles verdanke! Du hast mir im Verlauf der letzten Jahre nicht nur die Möglichkeit eröffnet, in die Welt der pädagogisch-psychologischen Forschung einzutauchen, sondern mich dabei auch in einer Art und Weise begleitet, die ich mir nicht anders hätte erträumen können. Du warst offen für meine Ideen, hast geduldig meine zahllosen Fragen beantwortet, dich wieder und wieder unermüdlich in meine Probleme hineingedacht, häufig genug den Überblick behalten und v.a. Gemeinsamkeiten gesehen (;-)), mir immer viel zugetraut und mich damit bestärkt und motiviert. Du hast bei jedem neuen Versuch im Publikationszirkus mitgefiebert, aber auch das Erholungspotential von Konferenzen aufgezeigt (;-)), du hast mich sogar in gestresstem Zustand gut ausgehalten, dich für mein Befinden interessiert, und mir in einer Arbeitswelt voll kurioser Anstellungsverhältnisse große Sicherheit vermittelt. Dabei habe ich nicht nur viel über die Themen der Pädagogischen Psychologie und ihre Erforschung gelernt, sondern konnte auch einen Einblick in dein Verständnis von guter Wissenschaft und deinen Umgang mit dem dazugehörigen Betrieb gewinnen. Außerdem begreife ich es als großes Glück, dass du nicht nur in wissenschaftlicher, sondern auch in menschlicher Hinsicht ein feines Vorbild bist. – Ganz zu schweigen von deinen Qualitäten als Elektriker, Umzugshelfer, Steuerberater, Brotkäufer und interessanter Gesprächspartner in diversen politischen und theologischen Fragen. ;-) Es ist ein Vorrecht, dich als Chef zu haben und ich bin ganz gespannt, mit welchen Projekten es nach Abschluss der Promotion dann weitergehen wird. Herzlichsten Dank für deine Unterstützung!

Die Arbeitsgruppe Pädagogische Psychologie bildet nicht nur eine Arbeitsumgebung, in der man sich wohlfühlt, sondern war besonders bei der Planung meines Projekts und der Entwicklung meiner Erhebungsinstrumente eine große Hilfe. Danke! Darüber hinaus gab es besondere Highlights wie die gemeinsamen Besuche von Fachtagungen mit meiner Kollegin Verena, aber ebenso auch alltägliche Freuden wie die ausgiebigen, häufig inspirierenden und jedenfalls entspannenden Mittagspausen mit Veronika und Franz. :-)

Jeder, der empirisch forscht und auf die Teilnahmebereitschaft von Anderen angewiesen ist, kennt das tiefe Gefühl der Dankbarkeit gegenüber der eigenen Stichprobe. Sich in dem sensiblen Bereich der Lehre für eine Beobachtung von außen zu öffnen, erforderte zum einen vielleicht etwas Mut, zum anderen aber vor allem die Bereitschaft, mir einen Teil der kostbaren eigenen Arbeitszeit und für die Erhebungen z.T. auch Veranstaltungszeit zur Verfügung zu stellen. Ohne dieses Entgegenkommen wäre diese Arbeit nicht möglich gewesen. Daher gilt mein besonderer Dank allen Lehrenden der Universität Kassel, die an meiner Studie teilgenommen haben.

Der Beobachtung und Einschätzung der Lehre in den untersuchten Lehrveranstaltungen kommt in dieser Studie eine zentrale Rolle zu. Die Güte meiner Analysen beruht ganz entscheidend darauf, wie die insgesamt sieben studentischen Hilfskräfte diese durchaus anspruchsvolle und

zeitintensive Aufgabe erfüllt haben. Und auch wenn es sich natürlich keineswegs um fehlerfreie Messungen handelt, so bin ich insgesamt doch recht zufrieden und dankbar für ihr Engagement.

Auf den letzten Metern haben jetzt doch noch viele mit angepackt, v.a. Heike, Hannah, Bea und Sammy. Dafür bin ich total dankbar! Es tut so gut, sich von Familie und Freunden bedacht, getragen und unterstützt zu wissen.

Zuletzt Gott. – In my father’s hand I am called to live; that's the safest place that life can give.

Appendix

Scales in Original Version (German)

A: Online Questionnaire for Teachers

Constructivist Teaching Beliefs: Konstruktivistische Lehr-Überzeugung

Skalenbezeichnung:	ST_konst
Anzahl der Items:	5
Antwortformat:	1 (trifft gar nicht zu) – 6 (trifft völlig zu)
N:	85
M (*SD*), Min, Max:	5.33 (0.63), 3.4, 6.0
Cronbach's α:	.68

Welche Überzeugungen haben Sie zur Hochschullehre? Welche Schwerpunkte setzen Sie und was ist Ihnen bei der Gestaltung Ihrer Lehre wichtig? Kreuzen Sie an, inwieweit die folgenden Aussagen auf Sie zutreffen!

Item		*M* (*SD*)	r_{it}
ST_konst_1	Es ist mein Anliegen, dass die Studierenden neue Perspektiven entdecken bzw. neue Denkweisen erlernen.	5.62 (0.69)	.54
ST_konst_2	Ich verstehe meine Lehre als Unterstützung der Studierenden bei der Entwicklung neuer Gedanken bzw. dem Aufbau neuen Wissens.	5.47 (0.67)	.47
ST_konst_3	In meinen Lehrveranstaltungen ermutige ich die Studierenden dazu, ihr bestehendes Wissen in Frage zu stellen und ggf. zu restrukturieren.	5.35 (0.98)	.60
ST_konst_4	Ich glaube, dass provokante Fragen und Hinweise auf inhaltliche Widersprüche sehr lernförderlich sein können.	5.40 (0.88)	.37
ST_konst_5	Ich lege bei meiner Lehrvorbereitung einen Schwerpunkt darauf, mir zu überlegen, wie ich die Studierenden zum Mitdenken bringen bzw. involvieren kann.	4.96 (1.18)	.33

Transmissive Teaching Beliefs: Transmissive Lehr-Überzeugung

Skalenbezeichnung:	ST_trans
Anzahl der Items:	5
Antwortformat:	1 (trifft gar nicht zu) – 6 (trifft völlig zu)
N:	81
M (*SD*), Min, Max:	4.48 (0.71), 2.2, 5.7
Cronbach's α:	.65

Welche Überzeugungen haben Sie zur Hochschullehre? Welche Schwerpunkte setzen Sie und was ist Ihnen bei der Gestaltung Ihrer Lehre wichtig? Kreuzen Sie an, inwieweit die folgenden Aussagen auf Sie zutreffen!			
Item		*M* (*SD*)	r_{it}
ST_trans_1	Als Dozent ist es meine zentrale Aufgabe, die Lehrinhalte gut zu präsentieren.	5.07 (1.07)	.55
ST_trans_2	Ich versuche, in meinen Lehrveranstaltungen möglichst viele Inhalte zu vermitteln, damit die Studierenden das erforderliche Wissen erwerben.	3.79 (1.12)	.25
ST_trans_3	Es ist wichtig, das Wissen möglichst strukturiert aufzubereiten, damit die Studierenden es gut aufnehmen können.	5.12 (0.99)	.43
ST_trans_4	Meine Aufgabe als Dozent ist es, die Studierenden mit dem Wissen zu versorgen, das sie zum Bestehen der Prüfung brauchen.	3.90 (1.45)	.53
ST_trans_5	Die Studierenden lernen neue Inhalte am besten, wenn ich sie ihnen gut erkläre.	4.80 (0.95)	.42

B: Entry Questionnaire for Students

Initial Student Interest: Interesse t_1

Skalenbezeichnung:	Int
Anzahl der Items:	7
Antwortformat:	1 (trifft gar nicht zu) – 6 (trifft völlig zu)
N:	4829
M (*SD*), Min, Max:	4.12 (1.02), 1.0, 6.0
Cronbach's α:	.89

Inwieweit interessiert Sie das Thema dieser Lehrveranstaltung?			
Item		*M* (*SD*)	r_{it}
Int_1	Viele der Themen, die in dieser Lehrveranstaltung behandelt werden sollen, finde ich sehr interessant.	4.25 (1.21)	.77
Int_2	Mit einigen der Themen würde ich mich gern intensiver beschäftigen.	3.91 (1.32)	.74
Int_3	Ich glaube, ich werde in dieser Veranstaltung Dinge lernen, die für mich wichtig sind.	4.33 (1.24)	.61
Int_4	Ich glaube, die Beschäftigung mit den Themen dieser Lehrveranstaltung wird mir Spaß machen.	3.90 (1.25)	.77
Int_5_in	Eigentlich interessiere ich mich gar nicht für die Themen dieser Lehrveranstaltung. (-)	4.87 (1.35)	.70
Int_6_in	In meiner Freizeit würde ich mich nie mit Themen dieser Lehrveranstaltung beschäftigen. (-)	3.99 (1.56)	.55
Int_7	Ich freue mich jetzt schon auf den nächsten Termin dieser Lehrveranstaltung.	3.58 (1.31)	.63

First Impression: Positiver Ersteindruck

Skalenbezeichnung:	positiv_Eindruck
Anzahl der Items:	5
Antwortformat:	1 (nein, gar nicht) – 6 (ja, sehr), -1 (weiß nicht)
N:	4428
M (*SD*), Min, Max:	4.83 (0.81), 1.0, 6.0
Cronbach's α:	.80

Ich glaube, der Dozent ist …			
Item		*M* (*SD*)	r_{it}
D_fair	fair,	5.05 (0.92)	.56
D_zugänglich	zugänglich und verständnisvoll,	4.81 (1.04)	.64
D_organisiert	gut organisiert,	5.01 (0.99)	.39
D_unterhaltsam	unterhaltsam,	4.66 (1.26)	.61
D_erklären	und kann gut erklären.	4.96 (1.00)	.71

C: Closing Questionnaire for Students

Structure: Strukturiertheit

Skalenbezeichnung:	L_Struk
Anzahl der Items:	5
Antwortformat:	1 (trifft gar nicht zu) – 6 (trifft völlig zu)
N:	2584
M (*SD*), Min, Max:	4.59 (0.93), 1.0, 6.0
Cronbach's α:	.86

Kreuzen Sie an, inwieweit die folgenden Aussagen Ihrer Ansicht nach auf diese Lehrveranstaltung zutreffen!			
Item		*M* (*SD*)	r_{it}
L_Struk_1	Die Lehrveranstaltung hatte einen logischen Aufbau.	4.70 (1.10)	.64
L_Struk_2	Wenn der Dozent etwas erklärt oder vorgetragen hat, war der rote Faden immer gut erkennbar.	4.55 (1.18)	.72
L_Struk_3_in	Die Struktur der Lehrveranstaltung war nicht so gut zu erkennen. (-)	4.76 (1.22)	.63
L_Struk_4	Ich wusste immer, „wo" (bei welchem inhaltlichen Punkt) der Dozent gerade ist.	4.31 (1.26)	.62
L_Struk_5	Die Beiträge des Dozenten waren gut strukturiert.	4.66 (1.08)	.77

Clarity: Verständlichkeit

Skalenbezeichnung:	L_Verst
Anzahl der Items:	2
Antwortformat:	1 (trifft gar nicht zu) – 6 (trifft völlig zu)
N:	2573
M (*SD*), Min, Max:	4.63 (1.06), 1.0, 6.0
Cronbach's α:	.86

Kreuzen Sie an, inwieweit die folgenden Aussagen Ihrer Ansicht nach auf diese Lehrveranstaltung zutreffen!

Item		*M* (*SD*)	r_{it}
L_Verst_1	Der Dozent hat die Lehrinhalte so erklärt, dass man sie gut verstehen konnte.	4.69 (1.12)	.75
L_Verst_2	Die Erklärungen/ Ausführungen des Dozenten waren immer gut verständlich.	4.58 (1.14)	.75

Enthusiasm: Enthusiasmus t_2

Skalenbezeichnung:	Enth_A
Anzahl der Items:	7
Antwortformat:	1 (trifft gar nicht zu) – 6 (trifft völlig zu)
N:	2557
M (*SD*), Min, Max:	5.01 (0.84), 1.0, 6.0
Cronbach's α:	.91

Jetzt geht es um Ihren Dozenten.

Item		*M* (*SD*)	r_{it}
Enth_L_1A	Der Dozent lehrt mit großem Enthusiasmus.	5.06 (1.11)	.77
Enth_F_1A	Wenn der Dozent sich mit Fragen seines Fachgebiets beschäftigt, ist er so richtig in seinem Element.	5.24 (0.93)	.70
Enth_FL_1A	Es bereitet dem Dozenten große Freude, den Studierenden sein Fachgebiet nahezubringen.	5.11 (1.02)	.83
Enth_F_2A	Der Dozent freut sich immer sehr, wenn sich jemand für sein Fachgebiet interessiert.	5.24 (0.92)	.65
Enth_L_5A	Beim Lehren blüht der Dozent so richtig auf.	4.87 (1.16)	.79
Enth_FL_3A	Der Dozent gibt sein Wissen zu seinem Fachgebiet gern weiter.	5.24 (0.87)	.74
Enth_L_6A	Dem Dozenten scheint das Konzipieren und Durchführen von Lehrveranstaltungen Spaß zu machen.	4.91 (1.11)	.71

Humour: Humor

Skalenbezeichnung:	D_Humor
Anzahl der Items:	2
Antwortformat:	1 (trifft gar nicht zu) – 6 (trifft völlig zu)
N:	2538
M (*SD*), Min, Max:	4.67 (1.27), 1.0, 6.0
Cronbach's α:	.86

Jetzt geht es um Ihren Dozenten.

Item		*M* (*SD*)	r_{it}
D_Unterh_1	Der Dozent gestaltet seine Lehre unterhaltsam.	4.52 (1.39)	.76
D_Unterh_2	Der Dozent hat Humor.	4.81 (1.32)	.76

Stimulation of Student Involvement: Förderung von Mitarbeit & kognitive Aktivierung

Skalenbezeichnung:	D_Mit_kognAk
Anzahl der Items:	5
Antwortformat:	1 (trifft gar nicht zu) – 6 (trifft völlig zu)
N:	2540
M (*SD*), Min, Max:	4.43 (1.03), 1.0, 6.0
Cronbach's α:	.88

Jetzt geht es um Ihren Dozenten.

Item		*M* (*SD*)	r_{it}
D_Mitarbeit_1	Der Dozent hat die Studierenden aktiv in die Lehrveranstaltung miteinbezogen.	4.54 (1.28)	.74
D_Mitarbeit_2	Es war dem Dozenten wichtig, dass die Studierenden mitarbeiten/ mitdenken.	4.81 (1.17)	.75
D_Fragen	Der Dozent hat den Studierenden viele Fragen gestellt.	4.01 (1.35)	.70
D_kognAk_1	Der Dozent hat gute (interessante/ herausfordernde) Fragen gestellt.	4.30 (1.26)	.71
D_kognAk_2	Der Dozent konnte die Studierenden zum Nachdenken bringen.	4.51 (1.21)	.66

Rapport: Guter Umgang mit Studierenden

Skalenbezeichnung:	D_UmS
Anzahl der Items:	7
Antwortformat:	1 (trifft gar nicht zu) – 6 (trifft völlig zu)
N:	2556
M (*SD*), Min, Max:	4.92 (0.86), 1.0, 6.0
Cronbach's α:	.89

Jetzt geht es um Ihren Dozenten.

Item		*M* (*SD*)	r_{it}
D_UmS_1	Der Dozent ist aufgeschlossen für Fragen und Probleme der Studierenden.	5.11 (1.07)	.76
D_UmS_2	Der Dozent ist offen für andere Meinungen.	4.83 (1.12)	.71
D_UmS_3	Man konnte dem Dozenten problemlos Fragen stellen.	5.23 (1.04)	.75
D_UmS_4	Der Dozent äußert bei guten studentischen Beiträgen seine Anerkennung.	4.73 (1.19)	.60
D_UmS_5	Der Dozent begegnet den Studierenden respektvoll.	5.32 (0.95)	.71
D_UmS_6	Der Dozent gab hilfreiches Feedback auf die Beiträge von Studierenden.	4.40 (1.23)	.59
D_UmS_7	Der Dozent ist den Studierenden sehr zugewandt.	4.87 (1.11)	.73

Cognitive Involvement

Skalenbezeichnung:	N_CogInv
Anzahl der Items:	6
Antwortformat:	1 (trifft gar nicht zu) – 6 (trifft völlig zu)
N:	2541
M (*SD*), Min, Max:	4.11 (0.98), 1.0, 6.0
Cronbach's α:	.81

Man kann in verschiedenen Lehrveranstaltungen unterschiedlich stark involviert sein. Kreuzen Sie an, was auf Sie in dieser Lehrveranstaltung am ehesten zutrifft!			
Item		*M* (*SD*)	r_{it}
N_Mitdenken_1	Während der Lehrveranstaltung habe ich eigentlich immer mitgedacht.	4.27 (1.12)	.66
N_Mitdenken_2_in	In dieser Lehrveranstaltung war ich meist unkonzentriert. (-)	4.66 (1.27)	.57
N_Mitdenken_3_in	Während der Veranstaltungszeit habe ich mich mit anderen Dingen beschäftigt oder mit Kommilitonen gequatscht. (-)	4.29 (1.32)	.44
N_Nachdenken_1	Ich habe mich außerhalb der Lehrveranstaltung mit anderen über die Inhalte unterhalten.	3.74 (1.56)	.54
N_Nachdenken_2	Manchmal habe ich noch nach der Lehrveranstaltung über die behandelten Inhalte nachgedacht.	4.02 (1.47)	.67
N_Nachdenken_3	Die Lehrveranstaltung hat mich zur kritischen Auseinandersetzung mit den Inhalten angeregt.	3.69 (1.41)	.61

Final Interest: Interesse t_2

Skalenbezeichnung:	Int_A
Anzahl der Items:	5
Antwortformat:	1 (trifft gar nicht zu) – 6 (trifft völlig zu)
N:	2617
M (*SD*), Min, Max:	4.10 (1.10), 1.0, 6.0
Cronbach's α:	.85

Wie interessant fanden Sie das Thema dieser Lehrveranstaltung?			
Item		*M* (*SD*)	r_{it}
Int_1A	Viele der Themen, die in dieser Lehrveranstaltung behandelt wurden, fand ich sehr interessant.	4.29 (1.31)	.80
Int_2A	Mit einigen der Themen würde ich mich gern weiter beschäftigen.	4.11 (1.38)	.77
Int_3A	Ich habe in dieser Veranstaltung Dinge gelernt, die für mich wichtig sind.	4.19 (1.34)	.72
Int_4A	Die Beschäftigung mit den Themen dieser Lehrveranstaltung hat mir Spaß gemacht.	4.01 (1.37)	.77
Int_6A_in	In meiner Freizeit würde ich mich nie mit Themen dieser Lehrveranstaltung beschäftigen. (-)	3.88 (1.55)	.31

Subjective Learning Achievement: Subjektiver Lernerfolg

Skalenbezeichnung:	Ziel_subLE
Anzahl der Items:	5
Antwortformat:	1 (trifft gar nicht zu) – 6 (trifft völlig zu)
N:	2523
M (*SD*), Min, Max:	4.46 (0.99), 1.0, 6.0
Cronbach's α:	.82

Wie würden Sie die Lehrveranstaltung rückblickend beurteilen?

Item		*M* (*SD*)	r_{it}
Sub_LE_1_in	Die Teilnahme an dieser Lehrveranstaltung war Zeitverschwendung. (-)	5.03 (1.32)	.64
Sub_LE_2_in	Wenn ich die Zeit für diese Lehrveranstaltung für Eigenstudium genutzt hätte, hätte ich mehr gelernt. (-)	4.57 (1.51)	.56
Sub_LE_3	Ich habe in dieser Lehrveranstaltung viel gelernt.	4.32 (1.29)	.74
Ziel_Wissen	Durch diese Lehrveranstaltung habe ich mein fachliches Wissen deutlich erweitert.	4.14 (1.23)	.67
Ziel_verstehen	Ich habe das Gefühl, die Inhalte gut verstanden zu haben.	4.30 (1.16)	.47

Development of Academic Competencies: Wissenschaftliche Kernkompetenzen

Skalenbezeichnung:	Ziel_wissKK
Anzahl der Items:	14
Antwortformat:	1 (trifft gar nicht zu) – 6 (trifft völlig zu)
N:	2519
M (*SD*), Min, Max:	3.33 (1.04), 1.0, 6.0
Cronbach's α:	.93

Wie würden Sie die Lehrveranstaltung rückblickend beurteilen?			
Item		*M* (*SD*)	r_{it}
Ziel_ Problem_1	In dieser Lehrveranstaltung habe ich mich darin verbessert, fachliche Problemstellungen zu bearbeiten.	3.36 (1.38)	.65
Ziel_ Problem_2	In dieser Lehrveranstaltung habe ich mich im problemlösenden Denken geübt.	3.24 (1.41)	.68
Ziel_analytD	Ich glaube, ich habe meine Fähigkeit, analytisch zu denken, durch diese Lehrveranstaltung verbessert.	3.29 (1.37)	.70
Ziel_begrMein_1	Ich habe in der Lehrveranstaltung gelernt, mir wissenschaftlich begründet eine Meinung zu einem Thema zu bilden und diese auch zu vertreten.	3.12 (1.42)	.75
Ziel_begrMein_2	Ich habe geübt, verschiedene Argumente abzuwägen und dann Stellung zu beziehen.	2.87 (1.40)	.74
Ziel_begrMein_3	Ich glaube, ich habe mich in dieser Lehrveranstaltung darin verbessert, zu einer Frage eine Position zu beziehen und diese gut zu begründen.	2.97 (1.41)	.76
Ziel_argument	Ich habe mich hier darin geübt, für oder gegen eine Aussage zu argumentieren.	2.69 (1.42)	.69
Ziel_Persp_1	Ich habe in dieser Lehrveranstaltung verschiedene Perspektiven und Theorien kennen gelernt, mit denen ein Phänomen auf unterschiedliche Art und Weise betrachtet werden kann.	3.60 (1.48)	.60
Ziel_Persp_2	Ich habe hier gelernt, etwas aus unterschiedlichen Perspektiven bzw. mit unterschiedlichen Theorien zu betrachten.	3.47 (1.43)	.73
Ziel_veknüpfen_1	Ich habe gelernt, verschiedene Themenbereiche miteinander zu verknüpfen und vernetzt zu denken.	3.72 (1.33)	.67
Ziel_veknüpfen_2	Die Lehrveranstaltung hat mir dabei geholfen, Verknüpfungen zwischen den behandelten Lehrinhalten und anderen Wissensgebieten oder Theorien herzustellen.	3.56 (1.36)	.69
Ziel_Nachdenken	Diese Lehrveranstaltung hat mich zum Nachdenken angeregt.	3.84 (1.45)	.63
Ziel_kritD	Ich wurde dazu angeregt, Sachverhalte, Thesen o.ä. kritisch zu hinterfragen.	3.40 (1.47)	.71
Ziel_Fragen	Diese Lehrveranstaltung hat bei mir weiterführende Fragen aufgeworfen.	3.17 (1.47)	.62

Revised Rating Form for Standardised Observation of Higher Education Teaching

Course: _____

Teacher: _____ Number of students (estimate): _____

Course format: _____ Date and time: _____

Observer: _____

Teaching methods		
	Time (in min)	Notes
▶ Organisational issues		
▶ Break during course time		
Teacher-guided methods (teacher leads the learning process):		
▶ Teacher talk		
▶ Guest speaker talk		
▶ Demonstrations (e.g. calculations, experiments)		
▶ Music or video		
▶		
Interactive methods (teacher interacts with students):		
▶ Teacher-moderated discussions		
▶ Text work		
▶		
Student-guided methods (students lead the learning process):		
▶ Student moderated discussions		
▶ Student presentation or micro-teaching		
▶		
Student-activating methods (all students are actively doing something without direct involvement of the teacher):		
▶ Individual work (e.g. reading, exercises)		
▶ Pair and group work (e.g. project work)		
▶ Activating methods involving all students together (e.g. games)		
▶		

Attention				
Students seem to be attentive.				
15 min in: _____ %	30 min in: _____ %	45 min in: _____ %	60 min in: _____ %	75 min in: _____ %
105 min in: _____ %	120 min in: _____ %	135 min in: _____ %	150 min in: _____ %	165 min in: _____ %

Media usage and visual aids

	Notes:
☐ Script, reader, book, text, handout	
☐ Projector (e.g. PowerPoint)	
☐ Black board, white board	
☐ Flip chart, posters	
☐ other (e.g. overhead projector)	

Quality of visual aids

	poor				very good	not applicable
Legibility	☐	☐	☐	☐	☐	☐
Illustration (pictures, charts)	☐	☐	☐	☐	☐	☐
Visual structuring (structuring slides, numbering, emphasis on important aspects)	☐	☐	☐	☐	☐	☐

Structure

	no, not at all				yes, notably or often	not applicable
Structure and aims: The structure of the session is presented and/ or aims of the session are clearly stated.	☐	☐	☐	☐	☐	☐
Connecting to prior knowledge: The teacher recaps content of past sessions and connects the new learning content to it; or makes a link to otherwise gained prior knowledge (homework, previous courses etc.).	☐	☐	☐	☐	☐	☐
Locating the topic: The teacher explains how the subject matter of the session fits in with the course or subject discipline.	☐	☐	☐	☐	☐	☐
Emphasising important content: Important parts of the content are pointed out, emphasised or repeated.	☐	☐	☐	☐	☐	☐
Summary: In the end and/ or in between, the learning content is summarised or conclusions are drawn.	☐	☐	☐	☐	☐	☐
Outlook: The teacher indicates what is coming up next (e.g. topics).	☐	☐	☐	☐	☐	☐

Rhetoric & Humour

Only applicable if teacher talks for a predefined minimum of time.

Elocution:						
The teacher talks …	no, not at all				yes, very much so	not applicable
coherently (≠ diffuse, disjointedly, illogically)	☐	☐	☐	☐	☐	☐
concisely (≠ unclear, circuitous)	☐	☐	☐	☐	☐	☐
eloquently (≠ inarticulately)	☐	☐	☐	☐	☐	☐
lively (≠ dully, monotonously)	☐	☐	☐	☐	☐	☐
comprehensibly (≠ too many foreign words, complicated phrasing)	☐	☐	☐	☐	☐	☐

Humour:						
	no, not at all				yes, very much so	not applicable
The teacher is funny. He uses puns, makes jokes, and tells anecdotes the students enjoy/ laugh about.	☐	☐	☐	☐	☐	☐

Verbal Illustration:						
The teacher …	no, not at all				yes, very much so	not applicable
connects the subject matter to everyday life experiences.	☐	☐	☐	☐	☐	☐
uses a lot of examples illustrating the subject matter.	☐	☐	☐	☐	☐	☐
uses a lot of comparisons and analogies.	☐	☐	☐	☐	☐	☐

Relevance:						
The teacher points out, that and how the subject matter is relevant for … (with explanation → higher rating)	no, not at all				yes, very much so	not applicable
the (final) exam	☐	☐	☐	☐	☐	☐
further study	☐	☐	☐	☐	☐	☐
the future job	☐	☐	☐	☐	☐	☐

Interaction

Rapport:

	no, not at all				yes, very much so	not applicable
The teacher treats the students with respect.	☐	☐	☐	☐	☐	☐
The teacher is friendly and appreciative towards the students.	☐	☐	☐	☐	☐	☐
The teacher is open for student suggestions, questions and other opinions.	☐	☐	☐	☐	☐	☐
The teacher takes the students questions and comments seriously and lets them finish.	☐	☐	☐	☐	☐	☐
The relationship between teacher and students seems to be tense. (-)	☐	☐	☐	☐	☐	☐

Teacher questions (≠ rhetoric questions, ≠ „Do you have any questions?“)

	number:
► organisational questions, unrelated to content (e.g. Did you attend course XY last semester?)	_____
► simple, content-related questions (e.g. Do you remember XY? – yes/no; What is XY? – word/sentence)	_____
► complex, open content-related questions (e.g. Why is XY? How can XY? – several sentences, e.g. explanation) and tasks, e.g. taking a stand, arguing for a position, substantiating an opinion, explaining a process (e.g. What do you think about XY and why? How would XY change if…?)	_____
► own remarks, answers and opinions (e.g. I myself think…, I do not agree with you in that…)	_____

Student contributions

	number:
► organisational questions and contributions (e.g. Which Moodle course? When is the exam?)	_____
► content-related questions, comprehension questions (e.g. How did X affect Y gain?)	_____
► critical questions or questions expanding on the content (e.g. But is this also true for XY?)	_____
► short contributions (single words or phrases), e.g. replies to teacher‘s question or task (e.g. X comes before Y.)	_____
► elaborate contributions (several sentences):	
► elaborate replies to teacher’s question or task, e.g. explanations (e.g. X = Y only if…; Regarding Z, X is preferable to Y, because…	_____
► personal ideas, experiences or opinions (e.g. In my family…; In my experience…; I like X better than Y…)	_____
► critical or analytical remarks and comments (e.g. I do not agree on XY, because…)	_____

Number of students participating:
Number of distinct students participating actively with verbal contributions. _____

Student participation						
	no, not at all				yes, very much so	not applicable
The students are engaged and participate actively.	☐	☐	☐	☐	☐	☐

Scientific standard (research focus, reflectivity, ambiguity, neutrality)						
	no, not at all				yes, very much so	not applicable
The teacher indicates the sources the presented learning content is based on (visually or verbally).	☐	☐	☐	☐	☐	☐
Contradictory research findings are presented (and discussed) or contradicting views are indicated.	☐	☐	☐	☐	☐	☐
The way of knowledge construction, the research methodology of the specific subject matter or the discipline as a whole is mentioned and discussed.	☐	☐	☐	☐	☐	☐
The fragility of findings, uncertainty of knowledge or multiple viewpoints on a matter are laid open.	☐	☐	☐	☐	☐	☐
The teacher advances a certain view without explicit mention or explanation. (-)	☐	☐	☐	☐	☐	☐
The teacher focuses strongly on certain topics that seem to be of personal importance or uses his own books only. (-)	☐	☐	☐	☐	☐	☐
The teacher leaves no room for other reasonable opinions. (-)	☐	☐	☐	☐	☐	☐
The teacher does not support the ideas presented with research, but seemingly largely presents own thoughts. (-)	☐	☐	☐	☐	☐	☐

Self-regulated learning			
Voluntary homework:			
	yes	no	not applicable
The teacher recommends the students to do some extra work on the subject matter (e.g. reading, research, experiments).	☐	☐	☐
The teacher gives literature recommendations for further reading into the topic.	☐	☐	☐
Obligatory homework: (Ask students beforehand about homework; if none → not applicable)			
	yes	no	not applicable
The teacher asks (and checks) whether the homework was done (and by whom).	☐	☐	☐
The teaching refers to or builds upon the homework.	☐	☐	☐
The teacher communicates clearly that he expects the students to do the homework or that he is unhappy about them not completing it.	☐	☐	☐

Personal (subjective) evaluation of the observer:	
On a scale from 1 (not at all) to 10 (very much):	
How much prior knowledge of the subject matter covered in this session did you have?	_____
How did you like the teacher?	_____
How would you rate the level of difficulty? How demanding was the subject matter?	_____

Was there anything particular about this session? _____

_____

_____

_____

_____

_____